The Rate-Distortion Function and Excess-Distortion Exponent of Sparse Regression Codes with Optimal Encoding

Ramji Venkataramanan, *Senior Member, IEEE,* and Sekhar Tatikonda, *Senior Member, IEEE*

***Abstract*—This paper studies the performance of sparse regression codes for lossy compression with the squared-error distortion criterion. In a sparse regression code, codewords are linear combinations of subsets of columns of a design matrix. It is shown that with minimum-distance encoding, sparse regression codes achieve the Shannon rate-distortion function for i.i.d. Gaussian sources $R^*(D)$ as well as the optimal excess-distortion exponent. This completes a previous result which showed that $R^*(D)$ and the optimal exponent were achievable for distortions below a certain threshold. The proof of the rate-distortion result is based on the second moment method, a popular technique to show that a non-negative random variable X is strictly positive with high probability. In our context, X is the number of codewords within target distortion D of the source sequence. We first identify the reason behind the failure of the standard second moment method for certain distortions, and illustrate the different failure modes via a stylized example. We then use a refinement of the second moment method to show that $R^*(D)$ is achievable for all distortion values. Finally, the refinement technique is applied to Suen's correlation inequality to prove the achievability of the optimal Gaussian excess-distortion exponent.**

***Index Terms*—Lossy compression, sparse superposition codes, rate-distortion function, Gaussian source, error exponent, second moment method, large deviations**

I. Introduction

DEVELOPING practical codes for lossy compression at rates approaching Shannon's rate-distortion bound has long been an important goal in information theory. A practical compression code requires a codebook with low storage complexity as well as encoding and decoding with low computational complexity. Sparse Superposition Codes or Sparse Regression Codes (SPARCs) are a recent class of codes introduced by Barron and Joseph, originally for communcation over the AWGN channel [1], [2]. They were subsequently used for lossy compression with the squared-error distortion criterion in [3], [4]. The codewords in a SPARC are linear combinations of columns of a design matrix $\mathbf{A}$. The storage complexity of the code is proportional to the size of the matrix, which is polynomial in the block length n. A computationally efficient encoder for compression with SPARCs was proposed in [4] and shown to achieve rates approaching the Shannon rate-distortion function for i.i.d. Gaussian sources.

R. Venkataramanan is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (e-mail: ramji.v@eng.cam.ac.uk).

S. Tatikonda is with the Department of Statistics and Data Science, Yale University, New Haven CT 06511, USA (e-mail: sekhar.tatikonda@yale.edu).

This work was partially supported by a Marie Curie Career Integration Grant (Grant Agreement Number 631489) and NSF Grant CCF-1217023. This paper was presented in part at the 2014 IEEE International Symposium on Information Theory.

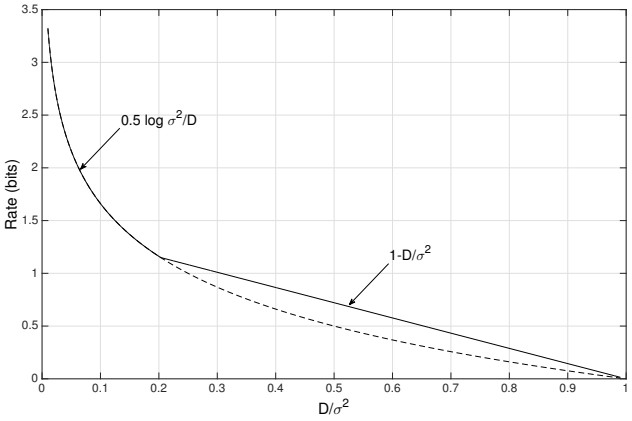

Fig. 1. The solid line shows the previous achievable rate $R_0(D)$, given in (1). The rate-distortion function $R^*(D)$ is shown in dashed lines. It coincides with $R_0(D)$ for $D/\sigma^2 \leq x^*$, where $x^* \approx 0.203$.

In this paper, we study the compression performance of SPARCs with the squared-error distortion criterion under optimal (minimum-distance) encoding. We show that for any ergodic source with variance σ^2, SPARCs with optimal encoding achieve a rate-distortion trade-off given by $R^*(D) := \frac{1}{2}\log\frac{\sigma^2}{D}$. Note that $R^*(D)$ is the optimal rate-distortion function for an i.i.d. Gaussian source with variance σ^2. The performance of SPARCs with optimal encoding was first studied in [3], where it was shown that for any distortion-level D, rates greater than

$$R_0(D) := \max\left\{\frac{1}{2}\log\frac{\sigma^2}{D},\ \left(1-\frac{D}{\sigma^2}\right)\right\} \tag{1}$$

are achievable with the optimal Gaussian excess-distortion exponent. The rate $R_0(D)$ in (1) is equal to $R^*(D)$ when $\frac{D}{\sigma^2} \leq x^*$, but is strictly larger than $R^*(D)$ when $\frac{D}{\sigma^2} > x^*$, where $x^* \approx 0.203$; see Fig. 1. In this paper, we complete the result of [3] by proving that sparse regression codes achieve the Gaussian rate-distortion function $R^*(D)$ for all distortions $D \in (0, \sigma^2)$. We also show that these codes attain the optimal excess-distortion exponent for i.i.d. Gaussian sources at all rates.

Though minimum-distance encoding is not practically feasible (indeed, the main motivation for sparse regression codes is that they enable low-complexity encoding and decoding), characterizing the rate-distortion function and excess-distortion exponent under optimal encoding establishes a benchmark to compare the performance of various computationally efficient

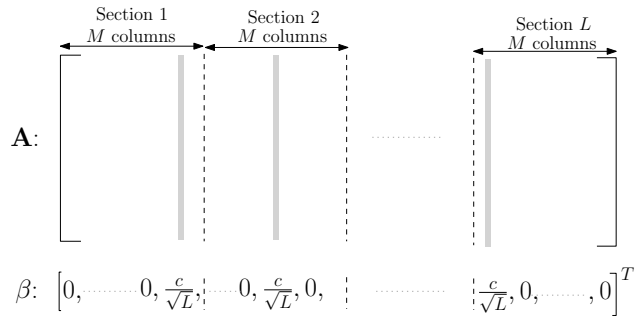

Fig. 2. $\mathbf{A}$ is an $n \times ML$ matrix and β is a $ML \times 1$ binary vector. The positions of the non-zeros in β correspond to the gray columns of $\mathbf{A}$ which combine to form the codeword $\mathbf{A}\beta$.

encoding schemes. Further, the results of this paper and [3] together show that SPARCs retain the good covering properties of the i.i.d. Gaussian random codebook, while having a compact representation in terms of a matrix whose size is a low-order polynomial in the block length.

Let us specify some notation before proceeding. Uppercase letters are used to denote random variables, and lowercase letters for their realizations. Bold-face letters are used to denote random vectors and matrices. All vectors have length n. The source sequence is $\mathbf{S} := (S_1, \ldots, S_n)$, and the reconstruction sequence is $\hat{\mathbf{S}} := (\hat{S}_1, \ldots, \hat{S}_n)$. $\|\mathbf{x}\|$ denotes the ℓ_2-norm of vector $\mathbf{x}$, and $|\mathbf{x}| = \frac{\|\mathbf{x}\|}{\sqrt{n}}$ is the normalized version. $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. Logarithms are with base e and rate is measured in nats, unless otherwise mentioned. The notation $a_n \sim b_n$ means that $\lim_{n\to\infty} \frac{1}{n} \log a_n = \lim_{n\to\infty} \frac{1}{n} \log b_n$, and w.h.p is used to abbreviate the phrase 'with high probability'. We will use $\kappa, \kappa_1, \kappa_2$ to denote generic positive constants whose exact value is not needed.

A. SPARCs with Optimal Encoding

A sparse regression code is defined in terms of a design matrix $\mathbf{A}$ of dimension $n \times ML$ whose entries are i.i.d. $\mathcal{N}(0, 1)$. Here n is the block length and M and L are integers whose values will be specified in terms of n and the rate R. As shown in Fig. 2, one can think of the matrix $\mathbf{A}$ as composed of L sections with M columns each. Each codeword is a linear combination of L columns, with one column from each section. Formally, a codeword can be expressed as $\mathbf{A}\beta$, where β is an $ML \times 1$ vector $(\beta_1, \ldots, \beta_{ML})$ with the following property: there is exactly one non-zero β_i for $1 \le i \le M$, one non-zero β_i for $M+1 \le i \le 2M$, and so forth. The non-zero values of β are all set equal to $\frac{c}{\sqrt{L}}$ where c is a constant that will be specified later. Denote the set of all β's that satisfy this property by $\mathcal{B}_{M,L}$.

Minimum-distance Encoder: This is defined by a mapping $g : \mathbb{R}^n \to \mathcal{B}_{M,L}$. Given the source sequence $\mathbf{S}$, the encoder determines the β that produces the codeword closest in Euclidean distance, i.e.,

$$g(\mathbf{S}) = \underset{\beta \in \mathcal{B}_{M,L}}{\operatorname{argmin}} \|\mathbf{S} - \mathbf{A}\beta\|.$$

Decoder: This is a mapping $h : \mathcal{B}_{M,L} \to \mathbb{R}^n$. On receiving $\beta \in \mathcal{B}_{M,L}$ from the encoder, the decoder produces reconstruction $h(\beta) = \mathbf{A}\beta$.

Since there are M columns in each of the L sections, the total number of codewords is M^L. To obtain a compression rate of R nats/sample, we therefore need

$$M^L = e^{nR}. \tag{2}$$

For our constructions, we choose $M = L^b$ for some $b > 1$ so that (2) implies

$$L \log L = \frac{nR}{b}. \tag{3}$$

Thus L is $\Theta(n/\log n)$, and the number of columns ML in the dictionary $\mathbf{A}$ is $\Theta\left((n/\log n)^{b+1}\right)$, a polynomial in n.

B. Overview of our Approach

To show that a rate R can be achieved at distortion-level D, we need to show that with high probability at least one of the e^{nR} choices for β satisfies

$$|\mathbf{S} - \mathbf{A}\beta|^2 \le D. \tag{4}$$

If β satisfies (4), we call it a *solution*.

Denoting the number of solutions by X, the goal is to show that $X > 0$ with high probability when $R > R^*(D)$. Note that X can be expressed as the sum of e^{nR} indicator random variables, where the ith indicator is 1 is $\beta(i)$ is a solution and zero otherwise, for $1 \le i \le e^{nR}$. Analyzing the probability $P(X > 0)$ is challenging because these indicator random variables are *dependent*: codewords $\mathbf{A}\beta(1)$ and $\mathbf{A}\beta(2)$ will be dependent if $\beta(1)$ and $\beta(2)$ share common non-zero terms. To handle the dependence, we use the second moment method (second MoM), a technique commonly used to prove existence ('achievability') results in random graphs and random constraint satisfaction problems [5]. In the setting of lossy compression, the second MoM was used in [6] to obtain the rate-distortion function of LDGM codes for binary symmetric sources with Hamming distortion.

For any non-negative random variable X, the second MoM [7] lower bounds the probability of the event $X > 0$ as[1]

$$P(X > 0) \ge \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}. \tag{5}$$

Therefore the second MoM succeeds if we can show that $(\mathbb{E}X)^2/\mathbb{E}[X^2] \to 1$ as $n \to \infty$. It was shown in [3] that the second MoM succeeds for $R > R_0(D)$, where $R_0(D)$ is defined in (1). In contrast, for $R^*(D) < R < R_0(D)$ it was found that $(\mathbb{E}X)^2/\mathbb{E}[X^2] \to 0$, so the second MoM fails. From this result in [3], it is not clear whether the gap from $R^*(D)$ is due to an inherent weakness of the sparse regression codebook, or if it is just a limitation of the second MoM as a proof technique. In this paper, we demonstrate that it is the latter, and refine the second MoM to prove that all rates greater than $R^*(D)$ are achievable.

Our refinement of the second MoM is inspired by the work of Coja-Oghlan and Zdeborová [8] on finding sharp thresholds

[1]The inequality (5) follows from the Cauchy-Schwarz inequality $(\mathbb{E}[XY])^2 \le \mathbb{E}X^2\,\mathbb{E}Y^2$ by substituting $Y = \mathbf{1}_{\{X>0\}}$.

for two-coloring of random hypergraphs. The high-level idea is as follows. The key ratio $(\mathbb{E}X)^2/\mathbb{E}[X^2]$ can be expressed as $(\mathbb{E}X)/\mathbb{E}[X(\beta)]$, where $X(\beta)$ denotes the total number of solutions conditioned on the event that a given β is a solution. (Recall that β is a solution if $|\mathbf{S} - \mathbf{A}\beta|^2 \leq D$.) Thus when the second MoM fails, i.e. the ratio goes to zero, we have a situation where the expected number of solutions is much smaller than the expected number of solutions *conditioned* on the event that β is a solution. This happens because for any $\mathbf{S}$, there are atypical realizations of the design matrix that yield a very large number of solutions. The total probability of these matrices is small enough that $\mathbb{E}X$ in not significantly affected by these realizations. However, conditioning on β being a solution increases the probability that the realized design matrix is one that yields an unusually large number of solutions. At low rates, the conditional probability of the design matrix being atypical is large enough to make $\mathbb{E}[X(\beta)] \gg \mathbb{E}X$, causing the second MoM to fail.[2]

The key to rectifying the second MoM failure is to show that $X(\beta) \approx \mathbb{E}X$ with high probability *although* $\mathbb{E}[X(\beta)] \gg \mathbb{E}X$. We then apply the second MoM to count just the 'good' solutions, i.e., solutions β for which $X(\beta) \approx \mathbb{E}X$. This succeeds, letting us conclude that $X > 0$ with high probability.

C. Related Work

As mentioned above, the second moment method was used in [6] to analyze the rate-distortion function of LDGM codes for binary symmetric sources with Hamming distortion. The idea of applying the second MoM to a random variable that counts just the 'good' solutions was recently used to obtain improved thresholds for problems such as random hypergraph 2-coloring [8], k-colorability of random graphs [9], and random k-SAT [10]. However, the key step of showing that a given solution is 'good' with high probability depends heavily on the geometry of the problem being considered. This step requires identifying a specific property of the random object being considered (e.g., SPARC design matrix, hypergraph, or boolean formula) that leads to a very large number of solutions in atypical realizations of the object. For example, in SPARC compression, the atypical realizations are design matrices with columns that are unusually well-aligned with the source sequence to be compressed; in random hypergraph 2-coloring, the atypical realizations are hypergraphs with an edge structure that allows an unusually large number of vertices to take on either color [8].

It is interesting to contrast the analysis of SPARC lossy compression with that of SPARC AWGN channel coding in [1]. The dependence structure of the SPARC codewords makes the analysis challenging in both problems, but the techniques required to analyze SPARC channel coding are very different from those used for the excess distortion analysis here. In the channel coding case, the authors use a modified union bound together with a novel bounding technique for the probability of pairwise error events [1, Lemmas 3,4] to establish that the error probability decays exponentially for all rates smaller than the channel capacity. In contrast, we use a refinement of the second moment method for the rate-distortion function, and Suen's correlation inequality to obtain the excess-distortion exponent.

Beyond the excess-distortion exponent, the *dispersion* is another quantity of interest in a lossy compression problem [11], [12]. For a fixed excess-distortion probability, the dispersion specifies how fast the rate can approach the rate-distortion function with growing block length. It was shown that for discrete memoryless and i.i.d. Gaussian sources, the optimal dispersion was equal to the inverse of the second derivative of the excess-distortion exponent. Given that SPARCs attain the optimal excess-distortion exponent, it would be interesting to explore if they also achieve the optimal dispersion for i.i.d. Gaussian sources with squared-error distortion.

The rest of the paper is organized as follows. The main results, specifying the rate-distortion function and the excess-distortion expoenent of SPARCs, are stated in Section II. In Section III, we set up the proof and show why the second MoM fails for $R < (1 - \frac{D}{\rho^2})$. As the proofs of the main theorems are technical, we motivate the main ideas with a stylized example in Section III-C. The proof of the main results are given in Section IV, with the proof of the main technical lemma given in Section V.

II. Main Results

The probability of excess distortion at distortion-level D of a rate-distortion code $\mathcal{C}_n$ with block length n and encoder and decoder mappings g, h is

$$P_e(\mathcal{C}_n, D) = P\left(|\mathbf{S} - h(g(\mathbf{S}))|^2 > D\right). \tag{6}$$

For a SPARC generated as described in Section I-A, the probability measure in (6) is with respect to the random source sequence $\mathbf{S}$ and the random design matrix $\mathbf{A}$.

A. Rate-Distortion Trade-off of SPARC

Definition 1. *A rate R is achievable at distortion level D if there exists a sequence of SPARCs $\{\mathcal{C}_n\}_{n=1,2,\ldots}$ such that $\lim_{n\to\infty} P_e(\mathcal{C}_n, D) = 0$ where for all n, $\mathcal{C}_n$ is a rate R code defined by an $n \times L_n M_n$ design matrix whose parameter L_n satisfies* (3) *with a fixed b and $M_n = L_n^b$.*

Theorem 1. *Let $\mathbf{S}$ be a drawn from an ergodic source with mean 0 and variance σ^2. For $D \in (0, \sigma^2)$, let $R^*(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$. Fix $R > R^*(D)$ and $b > b_{min}(\frac{\sigma^2}{D})$, where*

$$b_{min}(x) = \frac{20 R\, x^4}{\left(1+\frac{1}{x}\right)^2\left(1-\frac{1}{x}\right)\left[-1+\left(1+\frac{2\sqrt{x}}{(x-1)}\left(R-\frac{1}{2}(1-\frac{1}{x})\right)\right)^{1/2}\right]^2} \tag{7}$$

for $1 < x \leq e^{2R}$. Then there exists a sequence of rate R SPARCs $\{\mathcal{C}_n\}_{n=1,2\ldots}$ for which $\lim_{n\to\infty} P_e(\mathcal{C}_n, D) = 0$, where $\mathcal{C}_n$ is defined by an $n \times L_n M_n$ design matrix with $M_n = L_n^b$ and L_n determined by (3)*.*

[2] This is similar to the inspection paradox in renewal processes.

Remark: Though the theorem is valid for all $D \in (0, \sigma^2)$, it is most relevant for the case $\frac{D}{\sigma^2} > x^*$, where $x^* \approx 0.203$ is the solution to the equation

$$(1-x) + \frac{1}{2}\log x = 0.$$

For $\frac{D}{\sigma^2} \leq x^*$, [3, Theorem 1] already guarantees that the optimal rate-distortion function can be achieved, with a smaller value of b than that required by the theorem above.

B. Excess-distortion exponent of SPARC

The excess-distortion exponent at distortion-level D of a sequence of rate R codes $\{\mathcal{C}_n\}_{n=1,2,\ldots}$ is given by

$$r(D,R) = -\limsup_{n\to\infty} \frac{1}{n}\log P_e(\mathcal{C}_n, D), \tag{8}$$

where $P_e(\mathcal{C}_n, D)$ is defined in (6). The optimal excess-distortion exponent for a rate-distortion pair (R, D) is the supremum of the excess-distortion exponents over all sequences of codes with rate R, at distortion-level D.

The optimal excess-distortion exponent for discrete memoryless sources was obtained by Marton [13], and the result was extended to memoryless Gaussian sources by Ihara and Kubo [14].

Fact 1. *[14] For an i.i.d. Gaussian source distributed as $\mathcal{N}(0,\sigma^2)$ and squared-error distortion criterion, the optimal excess-distortion exponent at rate R and distortion-level D is*

$$r^*(D,R) = \begin{cases} \frac{1}{2}\left(\frac{a^2}{\sigma^2} - 1 - \log\frac{a^2}{\sigma^2}\right) & R > R^*(D) \\ 0 & R \leq R^*(D) \end{cases} \tag{9}$$

where $a^2 = De^{2R}$.

For $R > R^*(D)$, the exponent in (9) is the Kullback-Leibler divergence between two zero-mean Gaussians, distributed as $\mathcal{N}(0,a^2)$ and $\mathcal{N}(0,\sigma^2)$, respectively.

The next theorem characterizes the excess-distortion exponent performance of SPARCs.

Theorem 2. *Let $\mathbf{S}$ be drawn from an ergodic source with mean zero and variance σ^2. Let $D \in (0,\sigma^2)$, $R > \frac{1}{2}\log\frac{\sigma^2}{D}$, and $\gamma^2 \in (\sigma^2, De^{2R})$. Let*

$$b > \max\left\{2,\ \frac{7}{5}b_{min}\left(\gamma^2/D\right)\right\}, \tag{10}$$

where $b_{min}(.)$ is defined in (7). Then there exists a sequence of rate R SPARCs $\{C_n\}_{n=1,2\ldots}$, where $\mathcal{C}_n$ is defined by an $n \times L_n M_n$ design matrix with $M_n = L_n^b$ and L_n determined by (3), whose probability of excess distortion at distortion-level D can be bounded as follows for all sufficiently large n.

$$P_e(\mathcal{C}_n, D) \leq P(|\mathbf{S}|^2 \geq \gamma^2) + \exp\left(-\kappa n^{1+c}\right), \tag{11}$$

where κ, c are strictly positive constants.

Corollary 1. *Let $\mathbf{S}$ be drawn from an i.i.d. Gaussian source with mean zero and variance σ^2. Fix rate $R > \frac{1}{2}\log\frac{\sigma^2}{D}$, and let $a^2 = De^{2R}$. Fix any $\epsilon \in (0, a^2)$, and*

$$b > \max\left\{2,\ \frac{7}{5}b_{min}\left(\frac{a^2-\epsilon}{D}\right)\right\}. \tag{12}$$

There exists a sequence of rate R SPARCs with parameter b that achieves the excess-distortion exponent

$$\frac{1}{2}\left(\frac{a^2-\epsilon}{\sigma^2} - 1 - \log\frac{a^2-\epsilon}{\sigma^2}\right).$$

Consequently, the supremum of excess-distortion exponents achievable by SPARCs for i.i.d. Gaussian sources sources is equal to the optimal one, given by (9)*.*

Proof: From Theorem 2, we know that for any $\epsilon \in (0,\sigma^2)$, there exists a sequence of rate R SPARCs $\{C_n\}$ for which

$$P_e(\mathcal{C}_n, D) \leq P(|\mathbf{S}|^2 \geq a^2-\epsilon)\left(1 + \frac{\exp(-\kappa n^{1+c})}{P(|\mathbf{S}|^2 \geq a^2-\epsilon)}\right) \tag{13}$$

for sufficiently large n, as long as the parameter b satisfies (12). For $\mathbf{S}$ that is i.i.d. $\mathcal{N}(0,\sigma^2)$, Cramér's large deviation theorem [15] yields

$$\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n}\log P(|\mathbf{S}|^2 \geq a^2-\epsilon) \\ &= \frac{1}{2}\left(\frac{a^2-\epsilon}{\sigma^2} - 1 - \log\frac{a^2-\epsilon}{\sigma^2}\right) \end{aligned} \tag{14}$$

for $(a^2-\epsilon) > \sigma^2$. Thus $P(|\mathbf{S}|^2 \geq a^2-\epsilon)$ decays exponentially with n; in comparison $\exp(-\kappa n^{1+c})$ decays *faster* than exponentially with n. Therefore, from (13), the excess-distortion exponent satisfies

$$\begin{aligned} &\liminf_{n\to\infty} \frac{-1}{n}\log P_e(\mathcal{C}_n, D) \\ &\geq \liminf_{n\to\infty} \frac{-1}{n}\Big[\log P(|\mathbf{S}|^2 \geq a^2-\epsilon) \\ &\qquad + \log\left(1 + \frac{\exp(-\kappa n^{1+c})}{P(|\mathbf{S}|^2 \geq a^2-\epsilon)}\right)\Big] \\ &= \frac{1}{2}\left(\frac{a^2-\epsilon}{\sigma^2} - 1 - \log\frac{a^2-\epsilon}{\sigma^2}\right). \end{aligned} \tag{15}$$

Since $\epsilon > 0$ can be chosen arbitrarily small, the supremum of all achievable excess-distortion exponents is

$$\frac{1}{2}\left(\frac{a^2}{\sigma^2} - 1 - \log\frac{a^2}{\sigma^2}\right),$$

which is optimal from Fact 1. ∎

We remark that the function $b_{min}(x)$ is increasing in x. Therefore (12) implies that larger values of the design parameter b are required to achieve excess-distortion exponents closer to the optimal value (i.e., smaller values of ϵ in Corollary 1).

III. Inadequacy of the Direct Second MoM

A. First steps of the proof

Fix a rate $R > R^*(D)$, and b greater than the minimum value specified by the theorem. Note that $De^{2R} > \sigma^2$ since $R > \frac{1}{2}\log\frac{\sigma^2}{D}$. Let γ^2 be any number such that $\sigma^2 < \gamma^2 < De^{2R}$.

Code Construction: For each block length n, pick L as specified by (3) and $M = L^b$. Construct an $n \times ML$ design matrix $\mathbf{A}$ with entries drawn i.i.d. $\mathcal{N}(0,1)$. The codebook consists of all vectors $\mathbf{A}\beta$ such that $\beta \in \mathcal{B}_{M,L}$. The non-zero entries of β are all set equal to a value specified below.

Encoding and Decoding: If the source sequence $\mathbf{S}$ is such that $|\mathbf{S}|^2 \geq \gamma^2$, then the encoder declares an error. If $|\mathbf{S}|^2 \leq D$, then $\mathbf{S}$ can be trivially compressed to within distortion D using the all-zero codeword. The addition of this extra codeword to the codebook affects the rate in a negligible way.

If $|\mathbf{S}|^2 \in (D, \gamma^2)$, then $\mathbf{S}$ is compressed in two steps. First, quantize $|\mathbf{S}|^2$ with an n-level uniform scalar quantizer $Q(.)$ with support in the interval $(D, \gamma^2]$. For input $x \in (D, \gamma^2]$, if

$$x \in \left(D + \frac{(\gamma^2 - D)(i-1)}{n},\ D + \frac{(\gamma^2 - D)i}{n} \right],$$

for $i \in \{1, \ldots, n\}$, then the quantizer output is

$$Q(x) = D + \frac{(\gamma^2 - D)(i - \frac{1}{2})}{n}.$$

Conveying the scalar quantization index to the decoder (with an additional $\log n$ nats) allows us to adjust the codebook variance according to the norm of the observed source sequence.[3] The non-zero entries of β are each set to $\sqrt{(Q(|\mathbf{S}|^2) - D)/L}$ so that each SPARC codeword has variance $(Q(|\mathbf{S}|^2) - D)$. Define a "quantized-norm" version of $\mathbf{S}$ as

$$\tilde{\mathbf{S}} := \sqrt{\frac{Q(|\mathbf{S}|^2)}{|\mathbf{S}|^2}}\, \mathbf{S}. \tag{16}$$

Note that $|\tilde{\mathbf{S}}|^2 = Q(|\mathbf{S}|^2)$. We use the SPARC to compress $\tilde{\mathbf{S}}$. The encoder finds

$$\hat{\beta} := \underset{\beta \in \mathcal{B}_{M,L}}{\operatorname{argmin}} \ \|\tilde{\mathbf{S}} - \mathbf{A}\beta\|^2.$$

The decoder receives $\hat{\beta}$ and reconstructs $\hat{\mathbf{S}} = \mathbf{A}\hat{\beta}$. Note that for block length n, the total number of bits transmitted by encoder is $\log n + L \log M$, yielding an overall rate of $R + \frac{\log n}{n}$ nats/sample.

Error Analysis: For $\mathbf{S}$ such that $|\mathbf{S}|^2 \in (D, \gamma^2)$, the overall distortion can be bounded as

$$\begin{aligned} |\mathbf{S} - \mathbf{A}\hat{\beta}|^2 &= |\mathbf{S} - \tilde{\mathbf{S}} + \tilde{\mathbf{S}} - \mathbf{A}\hat{\beta}|^2 \\ &\leq |\mathbf{S} - \tilde{\mathbf{S}}|^2 + 2|\mathbf{S} - \tilde{\mathbf{S}}||\tilde{\mathbf{S}} - \mathbf{A}\hat{\beta}| + |\tilde{\mathbf{S}} - \mathbf{A}\hat{\beta}|^2 \\ &\leq \frac{\kappa_1}{n^2} + \frac{\kappa_2 |\tilde{\mathbf{S}} - \mathbf{A}\hat{\beta}|}{n} + |\tilde{\mathbf{S}} - \mathbf{A}\hat{\beta}|^2 \end{aligned} \tag{17}$$

for some positive constants κ_1, κ_2. The last inequality holds because the step-size of the scalar quantizer is $\frac{(\gamma^2 - D)}{n}$, and $|\mathbf{S}|^2 \in (D, \gamma^2)$.

Let $\mathcal{E}(\tilde{\mathbf{S}})$ be the event that the minimum of $|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2$ over $\beta \in \mathcal{B}_{M,L}$ is greater than D. The encoder declares an error if $\mathcal{E}(\tilde{\mathbf{S}})$ occurs. If $\mathcal{E}(\tilde{\mathbf{S}})$ *does not* occur, the overall distortion in (17) can be bounded as

$$|\mathbf{S} - \mathbf{A}\hat{\beta}|^2 \leq D + \frac{\kappa}{n}, \tag{18}$$

for some positive constant κ. The overall rate (including that of the scalar quantizer) is $R + \frac{\log n}{n}$.

Denoting the probability of excess distortion for this random code by $P_{e,n}$, we have

$$P_{e,n} \leq P(|\mathbf{S}|^2 \geq \gamma^2) + \max_{\rho^2 \in (D, \gamma^2)} P(\mathcal{E}(\tilde{\mathbf{S}}) \mid |\tilde{\mathbf{S}}|^2 = \rho^2). \tag{19}$$

As $\gamma^2 > \sigma^2$, the ergodicity of the source guarantees that

$$\lim_{n \to \infty} P(|\mathbf{S}|^2 \geq \gamma^2) = 0. \tag{20}$$

To bound the second term in (19), without loss of generality we can assume that the source sequence

$$\tilde{\mathbf{S}} = (\rho, \ldots, \rho).$$

This is because the codebook distribution is rotationally invariant, due to the i.i.d. $\mathcal{N}(0,1)$ design matrix $\mathbf{A}$. For any β, the entries of $\mathbf{A}\beta(i)$ i.i.d. $\mathcal{N}(0, \rho^2 - D)$. We enumerate the codewords as $\mathbf{A}\beta(i)$, where $\beta(i) \in \mathcal{B}_{M,L}$ for $i = 1, \ldots, e^{nR}$.

Define the indicator random variables

$$U_i(\tilde{\mathbf{S}}) = \begin{cases} 1 & \text{if } |\mathbf{A}\beta(i) - \tilde{\mathbf{S}}|^2 \leq D, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

We can then write

$$P(\mathcal{E}(\tilde{\mathbf{S}})) = P\left(\sum_{i=1}^{e^{nR}} U_i(\tilde{\mathbf{S}}) = 0 \right). \tag{22}$$

For a fixed $\tilde{\mathbf{S}}$, the $U_i(\tilde{\mathbf{S}})$'s are dependent. To see this, consider codewords $\hat{\mathbf{S}}(i), \hat{\mathbf{S}}(j)$ corresponding to the vectors $\beta(i), \beta(j) \in \mathcal{B}_{M,L}$, respectively. Recall that a vector in $\mathcal{B}_{M,L}$ is uniquely defined by the position of the non-zero value in each of its L sections. If $\beta(i)$ and $\beta(j)$ overlap in r of their non-zero positions, then the column sums forming codewords $\hat{\mathbf{S}}(i)$ and $\hat{\mathbf{S}}(j)$ will share r common terms, and consequently $U_i(\tilde{\mathbf{S}})$ and $U_j(\tilde{\mathbf{S}})$ will be dependent.

For brevity, we henceforth denote $U_i(\tilde{\mathbf{S}})$ by just U_i. Applying the second MoM with

$$X := \sum_{i=1}^{e^{nR}} U_i,$$

we have from (5)

$$P(X > 0) \geq \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]} \overset{(a)}{=} \frac{\mathbb{E}X}{\mathbb{E}[X \,|\, U_1 = 1]} \tag{23}$$

where (a) is obtained by expressing $\mathbb{E}[X^2]$ as follows.

$$\begin{aligned} \mathbb{E}[X^2] &= \mathbb{E}\left[X \sum_{i=1}^{e^{nR}} U_i \right] = \sum_{i=1}^{e^{nR}} \mathbb{E}[X U_i] \\ &= \sum_{i=1}^{e^{nR}} P(U_i = 1) \mathbb{E}[X | U_i = 1] \\ &= \mathbb{E}X \cdot \mathbb{E}[X \,|\, U_1 = 1]. \end{aligned} \tag{24}$$

The last equality in (24) holds because $\mathbb{E}X = \sum_{i=1}^{e^{nR}} P(U_i =$

[3]The scalar quantization step is only included to simplify the analysis. In fact, we could use the same codebook variance $(\gamma^2 - D)$ for all $\mathbf{S}$ that satisfy $|\mathbf{S}|^2 \leq (\gamma^2 - D)$, but this would make the forthcoming large deviations analysis quite cumbersome.

1), and due to the symmetry of the code construction. As $\mathbb{E}[X^2] \geq (\mathbb{E}X)^2$, (23) implies that $\mathbb{E}[X|\,U_1 = 1] \geq \mathbb{E}X$. Therefore, to show that $X > 0$ w.h.p, we need

$$\frac{\mathbb{E}[X|\,U_1 = 1]}{\mathbb{E}X} \to 1 \quad \text{as} \quad n \to \infty. \tag{25}$$

B. $\mathbb{E}X$ versus $\mathbb{E}[X|\,U_1 = 1]$

To compute $\mathbb{E}X$, we derive a general lemma specifying the probability that a randomly chosen i.i.d $\mathcal{N}(0, y)$ codeword is within distortion z of a source sequence $\mathbf{S}$ with $|\mathbf{S}|^2 = x$. This lemma will be used in other parts of the proof as well.

Lemma 1. *Let $\mathbf{S}$ be a vector with $|\mathbf{S}|^2 = x$. Let $\hat{\mathbf{S}}$ be an i.i.d. $\mathcal{N}(0, y)$ random vector that is independent of $\mathbf{S}$. Then for $x, y, z > 0$ and sufficiently large n, we have*

$$\frac{\kappa}{\sqrt{n}} e^{-nf(x,y,z)} \leq P\left(|\hat{\mathbf{S}} - \mathbf{S}|^2 \leq z\right) \leq e^{-nf(x,y,z)}, \tag{26}$$

where κ is a universal positive constant and for $x, y, z > 0$, the large-deviation rate function f is

$$f(x, y, z) = \begin{cases} \frac{x+z}{2y} - \frac{xz}{Ay} - \frac{A}{4y} - \frac{1}{2}\ln\frac{A}{2x} & \text{if } z \leq x + y, \\ 0 & \text{otherwise}, \end{cases} \tag{27}$$

and

$$A = \sqrt{y^2 + 4xz} - y. \tag{28}$$

Proof: We have

$$\begin{aligned} P\left(|\hat{\mathbf{S}} - \mathbf{S}|^2 \leq z\right) &= P\left(\frac{1}{n}\sum_{k=1}^{n}(\hat{S}_k - S_k)^2 \leq z\right) \\ &= P\left(\frac{1}{n}\sum_{k=1}^{n}(\hat{S}_k - \sqrt{x})^2 \leq z\right), \end{aligned} \tag{29}$$

where the last equality is due to the rotational invariance of the distribution of $\hat{\mathbf{S}}$, i.e., $\hat{\mathbf{S}}$ has the same joint distribution as $\mathbf{O}\hat{\mathbf{S}}$ for any orthogonal (rotation) matrix $\mathbf{O}$. In particular, we choose $\mathbf{O}$ to be the matrix that rotates $\mathbf{S}$ to the vector $(\sqrt{x}, \ldots, \sqrt{x})$, and note that $|\hat{\mathbf{S}} - \mathbf{S}|^2 = |\mathbf{O}\hat{\mathbf{S}} - \mathbf{O}\mathbf{S}|^2$. Then, using the strong version of Cramér's large deviation theorem due to Bahadur and Rao [15], [16], we have

$$\frac{\kappa}{\sqrt{n}} e^{-nI(x,y,z)} \leq P\left(\frac{1}{n}\sum_{k=1}^{n}(\hat{S}_k - x)^2 \leq z\right) \leq e^{-nI(x,y,z)}, \tag{30}$$

where the large-deviation rate function I is given by

$$I(x, y, z) = \sup_{\lambda \geq 0}\left\{\lambda z - \log \mathbb{E}e^{\lambda(\hat{S} - \sqrt{x})^2}\right\}. \tag{31}$$

The expectation on the RHS of (31) is computed with $\hat{S} \sim \mathcal{N}(0, y)$. Using standard calculations, we obtain

$$\log \mathbb{E}e^{\lambda(\hat{S} - \sqrt{x})^2} = \frac{\lambda x}{1 - 2y\lambda} - \frac{1}{2}\log(1 - 2y\lambda), \qquad \lambda < 2y. \tag{32}$$

Substituting the expression in (32) in (31) and maximizing over $\lambda \in [0, 2y)$ yields $I(x, y, z) = f(x, y, z)$, where f is given by (27). ∎

The expected number of solutions is given by

$$\mathbb{E}X = e^{nR}P(U_1 = 1) = e^{nR}P\left(|\mathbf{A}\beta(1) - \tilde{\mathbf{S}}|^2 \leq D\right). \tag{33}$$

Since $\tilde{\mathbf{S}} = (\rho, \rho, \ldots, \rho)$, and $\mathbf{A}\beta(1)$ is i.i.d. $\mathcal{N}(0, \rho^2 - D)$, applying Lemma 1 we obtain the bounds

$$\frac{\kappa}{\sqrt{n}} e^{nR} e^{-nf(\rho^2, \rho^2 - D, D)} \leq \mathbb{E}X \leq e^{nR} e^{-nf(\rho^2, \rho^2 - D, D)}, \tag{34}$$

Note that

$$f(\rho^2, \rho^2 - D, D) = \frac{1}{2}\log\frac{\rho^2}{D}. \tag{35}$$

Next consider $\mathbb{E}[X|\,U_1 = 1]$. If $\beta(i)$ and $\beta(j)$ overlap in r of their non-zero positions, the column sums forming codewords $\hat{\mathbf{S}}(i)$ and $\hat{\mathbf{S}}(j)$ will share r common terms. Therefore,

$$\begin{aligned} \mathbb{E}[X|\,U_1 = 1] &= \sum_{i=1}^{e^{nR}} P(U_i = 1|\,U_1 = 1) \\ &= \sum_{i=1}^{e^{nR}} \frac{P(U_i = 1,\, U_1 = 1)}{P(U_1 = 1)} \\ &\overset{(a)}{=} \sum_{r=0}^{L} \binom{L}{r}(M-1)^{L-r} \frac{P(U_2 = U_1 = 1|\,\mathcal{F}_{12}(r))}{P(U_1 = 1)} \end{aligned} \tag{36}$$

where $\mathcal{F}_{12}(r)$ is the event that the codewords corresponding to U_1 and U_2 share r common terms. In (36), (a) holds because for each codeword $\hat{\mathbf{S}}(i)$, there are a total of $\binom{L}{r}(M-1)^{L-r}$ codewords which share exactly r common terms with $\hat{\mathbf{S}}(i)$, for $0 \leq r \leq L$. From (36) and (33), we obtain

$$\begin{aligned} &\frac{\mathbb{E}[X|\,U_1 = 1]}{\mathbb{E}X} \\ &= \sum_{r=0}^{L} \binom{L}{r}(M-1)^{L-r} \frac{P(U_2 = U_1 = 1|\,\mathcal{F}_{12}(r))}{e^{nR}\,(P(U_1 = 1))^2} \\ &\overset{(a)}{\sim} 1 + \sum_{\alpha = \frac{1}{L}, \ldots, \frac{L}{L}} \binom{L}{L\alpha} \frac{P(U_2 = U_1 = 1|\,\mathcal{F}_{12}(\alpha))}{M^{L\alpha}\,(P(U_1 = 1))^2} \\ &\overset{(b)}{=} 1 + \sum_{\alpha = \frac{1}{L}, \ldots, \frac{L}{L}} e^{n\Delta_\alpha} \end{aligned} \tag{37}$$

where (a) is obtained by substituting $\alpha = \frac{r}{L}$ and $e^{nR} = M^L$. The notation $x_L \sim y_L$ means that $x_L/y_L \to 1$ as $L \to \infty$. The equality (b) is from [3, Appendix A], where it was also shown that

$$\Delta_\alpha \leq \frac{\kappa}{L} + \frac{R}{b}\min\{\alpha, \bar{\alpha}, \tfrac{\log 2}{\log L}\} - h(\alpha) \tag{38}$$

where

$$h(\alpha) := \alpha R - \frac{1}{2}\log\left(\frac{1 + \alpha}{1 - \alpha(1 - \frac{2D}{\rho^2})}\right). \tag{39}$$

The inequality in (38) is asymptotically tight [3]. The term $e^{n\Delta_\alpha}$ in (37) may be interpreted as follows. Conditioned on $\beta(1)$ being a solution, the expected number of solutions that share αL common terms with $\beta(1)$ is $\sim e^{n\Delta_\alpha}\mathbb{E}X$. Recall that we require the left side of (37) to tend to 1 as $n \to \infty$. Therefore, we need $\Delta_\alpha < 0$ for $\alpha = \frac{1}{L}, \ldots, \frac{L}{L}$. From (38), we need $h(\alpha)$ to be positive in order to guarantee that $\Delta_\alpha < 0$.

However, when $R < (1 - \frac{D}{\rho^2})$, it can be verified that $h(\alpha) < 0$ for $\alpha \in (0, \alpha^*)$ where $\alpha^* \in (0,1)$ is the solution to $h(\alpha) = 0$. Thus Δ_α is *positive* for $\alpha \in (0, \alpha^*)$ when $\frac{1}{2}\log\frac{\rho^2}{D} \leq R \leq (1 - \frac{D}{\rho^2})$. Consequently, (37) implies that

$$\frac{\mathbb{E}[X|\, U_1 = 1]}{\mathbb{E}X} \sim \sum_\alpha e^{n\Delta_\alpha} \to \infty \;\text{ as }\; n \to \infty, \tag{40}$$

and the second MoM fails.

C. A Stylized Example

Before describing how to rectify the second MoM failure in the SPARC setting, we present a simple example to give intuition about the failure modes of the second MoM. The proofs in the next two sections do not rely on the discussion here.

Consider a sequence of generic random structures (e.g., a sequence of random graphs or SPARC design matrices) denoted by R_n, $n \geq 1$. Suppose that for each n, the realization of R_n belongs to one of two categories: a category C_1 structure which has which has e^n solutions, or a category C_2 structure which has e^{2n} solutions. In the case of SPARC, a solution is a codeword that is within the target distortion. Let the probabilities of R_n being of each category be

$$P(R_n \in \mathsf{C}_1) = 1 - e^{-np}, \quad P(R_n \in \mathsf{C}_2) = e^{-np}, \tag{41}$$

where $p > 0$ is a constant. Regardless of the realization, we note that R_n always has at least e^n solutions.

We now examine whether the second MoM can guarantee the existence of a solution for this problem as $n \to \infty$. The number of solutions X can be expressed as a sum of indicator random variables:

$$X = \sum_{i=1}^{N} U_i,$$

where $U_i = 1$ if configuration i is a solution, and N is the total number of configurations. (In the SPARC context, a configuration is a codeword.) We assume that the configurations are symmetric (as in the SPARC set-up), so that each one has equal probability of being a solution, i.e.,

$$P(U_i = 1 \mid R_n \in \mathsf{C}_1) = \frac{e^n}{N}, \quad P(U_i = 1 \mid R_n \in \mathsf{C}_2) = \frac{e^{2n}}{N}. \tag{42}$$

Due to symmetry, the second moment ratio can be expressed as

$$\frac{\mathbb{E}X^2}{(\mathbb{E}X)^2} = \frac{\mathbb{E}[X \mid U_1 = 1]}{\mathbb{E}X} = \frac{\mathbb{E}[X \mid U_1 = 1]}{(1 - e^{-np})e^n + e^{-np}e^{2n}}. \tag{43}$$

The conditional expectation in the numerator can be computed as follows.

$$\begin{aligned}
\mathbb{E}[X|U_1 = 1] &= P(R_n \in \mathsf{C}_1 \mid U_1 = 1)\mathbb{E}[X|U_1 = 1, \mathsf{C}_1] \\
&\quad + P(R_n \in \mathsf{C}_2 \mid U_1 = 1)\mathbb{E}[X|U_1 = 1, \mathsf{C}_2] \\
&\overset{(a)}{=} \frac{(1 - e^{-np})(e^n/N)}{(1 - e^{-np})(e^n/N) + e^{-np}(e^{2n}/N)}\, e^n \\
&+ \frac{e^{-np}(e^{2n}/N)}{(1 - e^{-np})(e^n/N) + e^{-np}(e^{2n}/N)}\, e^{2n} \\
&= \frac{(1 - e^{-np})e^{2n} + e^{n(4-p)}}{(1 - e^{-np})e^n + e^{n(2-p)}},
\end{aligned} \tag{44}$$

where (a) is obtained by using Bayes' rule to compute $P(R_n \in \mathsf{C}_1 \mid U_1 = 1)$. The second MoM ratio in (43) therefore equals

$$\frac{\mathbb{E}X^2}{(\mathbb{E}X)^2} = \frac{\mathbb{E}[X \mid U_1 = 1]}{\mathbb{E}X} = \frac{(1 - e^{-np})e^{2n} + e^{n(4-p)}}{[(1 - e^{-np})e^n + e^{n(2-p)}]^2}. \tag{45}$$

We examine the behavior of the ratio above as $n \to \infty$ for different values of p.

Case 1**:** $p \geq 2$. The dominant term in both the numerator and the denominator of (45) is e^{2n}, and we get

$$\frac{\mathbb{E}[X \mid U_1 = 1]}{\mathbb{E}X} \to 1 \text{ as } n \to \infty, \tag{46}$$

and the second MoM succeeds.

Case 2**:** $1 < p \leq 2$. The dominant term in the numerator is $e^{n(4-p)}$, while the dominant term in the denominator is e^{2n}. Hence

$$\frac{\mathbb{E}[X \mid U_1 = 1]}{\mathbb{E}X} = \frac{e^{n(4-p)}}{e^{2n}}(1 + o(1)) \sim e^{n(2-p)} \overset{n\to\infty}{\longrightarrow} \infty. \tag{47}$$

Case 3**:** $0 < p \leq 1$. The dominant term in the numerator is $e^{n(4-p)}$, while the dominant term in the denominator is $e^{n(4-2p)}$. Hence

$$\frac{\mathbb{E}[X \mid U_1 = 1]}{\mathbb{E}X} = \frac{e^{n(4-p)}}{e^{n(4-2p)}}(1 + o(1)) \sim e^{np} \overset{n\to\infty}{\longrightarrow} \infty. \tag{48}$$

Thus in both Case 2 and Case 3, the second MoM fails because the expected number of solutions conditioned on a solution $(U_1 = 1)$ is exponentially larger than the unconditional expected value. However, there is an important distinction between the two cases, which allows us to fix the failure of the second MoM in Case 2 but not in Case 3.

Consider the conditional distribution of the number of solutions given $U_1 = 1$. From the calculation in (44), we have

$$\begin{aligned}
P(X = e^n \mid U_1 = 1) &= P(R_n \in \mathsf{C}_1 \mid U_1 = 1) \\
&= \frac{(1 - e^{-np})e^n}{(1 - e^{-np})e^n + e^{n(2-p)}}, \\
P(X = e^{2n} \mid U_1 = 1) &= P(R_n \in \mathsf{C}_2 \mid U_1 = 1) \\
&= \frac{e^{n(2-p)}}{(1 - e^{-np})e^n + e^{n(2-p)}}.
\end{aligned} \tag{49}$$

When $1 < p \leq 2$, the first term in the denominator of the

RHS dominates, and the conditional distribution of X is

$$\begin{aligned} P(X = e^n \mid U_1 = 1) &= 1 - e^{-n(p-1)}(1 + o(1)), \\ P(X = e^{2n} \mid U_1) &= e^{-n(p-1)}(1 + o(1)). \end{aligned} \tag{50}$$

Thus the conditional probability of a realization R_n being category C_1 given $U_1 = 1$ is slightly smaller than the unconditional probability, which is $1-e^{-np}$. However, conditioned on $U_1 = 1$, a realization R_n is still extremely likely to have come from category C_1, i.e., have e^n solutions. Therefore, when $1 < p \leq 2$, conditioning on a solution *does not* change the nature of the 'typical' or 'high-probability' realization. This makes it possible to fix the failure of the second MoM in this case. The idea is to define a new random variable X' which counts the number of solutions coming from typical realizations, i.e., only category C_1 structures. The second MoM is then applied to X' to show that is strictly positive with high probability.

When $p < 1$, conditioning on a solution completely changes the distribution of X. The dominant term in the denominator of the RHS in (49) is $e^{n(2-p)}$, so the conditional distribution of X is

$$\begin{aligned} P(X = e^n \mid U_1 = 1) &= e^{-n(1-p)}(1 + o(1)), \\ P(X = e^{2n} \mid U_1) &= 1 - e^{-n(1-p)}(1 + o(1)). \end{aligned} \tag{51}$$

Thus, conditioned on a solution, a typical realization of R_n belongs to category C_2, i.e., has e^{2n} solutions. On the other hand, if we draw from the unconditional distribution of R_n in (41), a typical realization has e^n solutions. In this case, the second moment method cannot be fixed by counting only the solutions from realizations of category C_1, because the total conditional probability of such realizations is very small. This is the analog of the "condensation phase" that is found in problems such as random hypergraph coloring [8]. In this phase, although solutions may exist, even an enhanced second MoM does not prove their existence.

Fortunately, there is no condensation phase in the SPARC compression problem. Despite the failure of the direct second MoM, we prove (Lemma 2) that conditioning on a solution does not significantly alter the total number of solutions for a very large fraction of design matrices. Analogous to Case 2 above, we can apply the second MoM to a new random variable that counts only the solutions coming from typical realizations of the design matrix. This yields the desired result that solutions exist for all rates $R < R^*(D)$.

IV. Proofs of Main Results

A. Proof of Theorem 1

The code parameters, encoding and decoding are as described in Section III-A. We build on the proof set-up of Section III-B. Given that $\beta \in \mathcal{B}_{M,L}$ is a solution, for $\alpha = 0, \frac{1}{L}, \ldots, \frac{L}{L}$ define $X_\alpha(\beta)$ to be the number of solutions that share αL non-zero terms with β. The *total* number of solutions given that β is a solution is

$$X(\beta) = \sum_{\alpha=0,\frac{1}{L},\ldots,\frac{L}{L}} X_\alpha(\beta) \tag{52}$$

Using this notation, we have

$$\begin{aligned} \frac{\mathbb{E}[X|\, U_1 = 1]}{\mathbb{E}X} &\overset{(a)}{=} \frac{\mathbb{E}[X(\beta)]}{\mathbb{E}X} \\ &= \sum_{\alpha=0,\frac{1}{L},\ldots,\frac{L}{L}} \frac{\mathbb{E}[X_\alpha(\beta)]}{\mathbb{E}X} \overset{(b)}{\sim} 1 + \sum_{\alpha=\frac{1}{L},\ldots,\frac{L}{L}} e^{n\Delta_\alpha}, \end{aligned} \tag{53}$$

where (a) holds because the symmetry of the code construction allows us to condition on a generic $\beta \in \mathcal{B}_{M,L}$ being a solution; (b) follows from (37). Note that $\mathbb{E}[X_\alpha(\beta)]$ and $\mathbb{E}[X(\beta)]$ are expectations evaluated with the *conditional* distribution over the space of design matrices given that β is a solution.

The key ingredient in the proof is the following lemma, which shows that $X_\alpha(\beta)$ is much smaller than $\mathbb{E}X$ w.h.p $\forall \alpha \in \{\frac{1}{L}, \ldots, \frac{L}{L}\}$. In particular, $X_\alpha(\beta) \ll \mathbb{E}X$ *even* for α for which

$$\frac{\mathbb{E}[X_\alpha(\beta)]}{\mathbb{E}X} \sim e^{n\Delta_\alpha} \to \infty \quad \text{as } n \to \infty.$$

Lemma 2. *Let $R > \frac{1}{2}\log\frac{\rho^2}{D}$. If $\beta \in \mathcal{B}_{M,L}$ is a solution, then for sufficiently large L*

$$P\left(X_\alpha(\beta) \leq L^{-3/2}\,\mathbb{E}X, \;\; \textit{for } \tfrac{1}{L} \leq \alpha \leq \tfrac{L-1}{L}\right) \geq 1 - \eta \tag{54}$$

where

$$\eta = L^{-2.5\left(\frac{b}{b_{min}(\rho^2/D)}-1\right)}. \tag{55}$$

The function $b_{min}(.)$ is defined in (7)*.*

Proof. The proof of the lemma is given in Section V.

The probability measure in Lemma 2 is the conditional distribution on the space of design matrices $\mathbf{A}$ given that β is a solution.

Definition 2. *For $\epsilon > 0$, call a solution β "ϵ-good" if*

$$\sum_{\alpha=\frac{1}{L},\ldots,\frac{L}{L}} X_\alpha(\beta) < \epsilon\,\mathbb{E}X. \tag{56}$$

Since we have fixed $\tilde{\mathbf{S}} = (\rho, \ldots, \rho)$, whether a solution β is ϵ-good or not is determined by the design matrix. Lemma 2 guarantees that w.h.p any solution β will be ϵ-good, i.e., if β is a solution, w.h.p the design matrix is such that the number of solutions sharing any common terms with β is less $\epsilon\mathbb{E}[X]$.

The key to proving Theorem 1 is to apply the second MoM only to ϵ-good solutions. Fix $\epsilon = L^{-0.5}$. For $i = 1, \ldots, e^{nR}$, define the indicator random variables

$$V_i = \begin{cases} 1 & \text{if } |\mathbf{A}\beta(i) - \tilde{\mathbf{S}}|^2 \leq D \textit{ and } \beta(i) \text{ is } \epsilon\text{-good}, \\ 0 & \text{otherwise.} \end{cases} \tag{57}$$

The number of ϵ-good solutions, denoted by X_g, is given by

$$X_g = V_1 + V_2 + \ldots + V_{e^{nR}}. \tag{58}$$

We will apply the second MoM to X_g to show that $P(X_g > 0) \to 1$ as $n \to \infty$. We have

$$P(X_g > 0) \geq \frac{(\mathbb{E}X_g)^2}{\mathbb{E}[X_g^2]} = \frac{\mathbb{E}X_g}{\mathbb{E}[X_g|\, V_1 = 1]} \tag{59}$$

where the second equality is obtained by writing $\mathbb{E}[X_g^2] = (\mathbb{E}X_g)\mathbb{E}[X_g|\, V_1 = 1]$, similar to (24).

Lemma 3. *a)* $\mathbb{E}X_g \geq (1-\eta)\mathbb{E}X$, *where* η *is defined in* (55). *b)* $\mathbb{E}[X_g|\, V_1 = 1] \leq (1 + L^{-0.5})\mathbb{E}X$.

Proof: Due to the symmetry of the code construction, we have

$$\begin{aligned}\mathbb{E}X_g &= e^{nR}P(V_1 = 1) \stackrel{(a)}{=} e^{nR}P(U_1 = 1)P(V_1 = 1|U_1 = 1)\\ &= \mathbb{E}X \cdot P(\beta(1) \text{ is } \epsilon\text{-good } \mid \beta(1) \text{ is a solution}).\end{aligned} \tag{60}$$

In (60), (a) follows from the definitions of V_i in (57) and U_i in (21). Given that $\beta(1)$ is a solution, Lemma 2 shows that

$$\sum_{\alpha=\frac{1}{L},\ldots,\frac{L}{L}} X_\alpha(\beta(1)) < (\mathbb{E}X)L^{-0.5}. \tag{61}$$

with probability at least $1-\eta$. As $\epsilon = L^{-0.5}$, $\beta(1)$ is ϵ-good according to Definition 2 if (61) is satisfied. Thus $\mathbb{E}X_g$ in (60) can be lower bounded as

$$\mathbb{E}X_g \geq (1-\eta)\mathbb{E}X. \tag{62}$$

For part (b), first observe that the total number of solutions X is an upper bound for the number of ϵ-good solutions X_g. Therefore

$$\mathbb{E}[X_g|\, V_1 = 1] \leq \mathbb{E}[X|\, V_1 = 1]. \tag{63}$$

Given that $\beta(1)$ is an ϵ-good solution, the expected number of solutions can be expressed as

$$\begin{aligned}&\mathbb{E}[X|\, V_1 = 1]\\ &= \mathbb{E}[X_0(\beta(1)) \mid V_1 = 1] + \mathbb{E}[\sum_{\alpha=\frac{1}{L},\ldots,\frac{L}{L}} X_\alpha(\beta(1)) \mid V_1 = 1].\end{aligned} \tag{64}$$

There are $(M-1)^L$ codewords that share *no* common terms with $\beta(1)$. Each of these codewords is independent of $\beta(1)$, and thus independent of the event $V_1 = 1$.

$$\begin{aligned}&\mathbb{E}[X_0(\beta(1)) \mid V_1 = 1] = \mathbb{E}[X_0(\beta(1))]\\ &= (M-1)^L\, P(|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 \leq D)\\ &\leq M^L\, P(|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 \leq D)\\ &= \mathbb{E}X.\end{aligned} \tag{65}$$

Next, note that conditioned on $\beta(1)$ being an ϵ-good solution (i.e., $V_1 = 1$),

$$\sum_{\alpha=\frac{1}{L},\ldots,\frac{L}{L}} X_\alpha(\beta(1)) < \epsilon\, \mathbb{E}X \tag{66}$$

with certainty. This follows from the definition of ϵ-good in (56). Using (65) and (66) in (64), we conclude that

$$\mathbb{E}[X|\, V_1 = 1] < (1+\epsilon)\mathbb{E}X. \tag{67}$$

Combining (67) with (63) completes the proof of Lemma 3. ∎

Using Lemma 3 in (59), we obtain

$$\begin{aligned}P(X_g > 0) &\geq \frac{\mathbb{E}X_g}{\mathbb{E}[X_g|\, V_1 = 1]} \geq \frac{(1-\eta)}{1+\epsilon}\\ &= \frac{1 - L^{-2.5(\frac{b}{b_{min}(\rho^2/D)}-1)}}{1 + L^{-1/2}},\end{aligned} \tag{68}$$

where the last equality is obtained by using the definition of η in (55) and $\epsilon = L^{-0.5}$. Hence the probability of the existence of at least one good solution goes to 1 as $L \to \infty$. Thus we have shown that for any $\rho^2 \in (D, \gamma^2)$, the quantity

$$P(\mathcal{E}(\tilde{\mathbf{S}}) \mid |\tilde{\mathbf{S}}|^2 = \rho^2)$$

in (19) tends to zero whenever $R > \frac{1}{2}\log\frac{\rho^2}{D}$ and $b > b_{min}(\frac{\rho^2}{D})$. Combining this with (18)–(20),we conclude that that the probability that

$$|\mathbf{S} - \mathbf{A}\hat{\beta}|^2 \leq D + \frac{\kappa}{n}$$

goes to one as $n \to \infty$. As $\gamma^2 > \sigma^2$ can be chosen arbitrarily close to σ^2, the proof of Theorem 1 is complete.

B. Proof of Theorem 2

The code construction is as described in Section III-A, with the parameter b now chosen to satisfy (10). Recall the definition of an ϵ-good solution in Definition 2. We follow the set-up of Section IV-A and count the number of ϵ-good solutions, for an appropriately defined ϵ. As before, we want an upper bound for the probability of the event $X_g = 0$, where the number of ϵ-good solutions X_g is defined in (58).

Theorem 2 is obtained using Suen's correlation inequality to upper bound on the probability of the event $X_g = 0$. Suen's inequality yields a sharper upper bound than the second MoM. We use it to prove that the probability of $X_g = 0$ decays *super-exponentially* in L. In comparison, the second MoM only guarantees a polynomial decay.

We begin with some definitions required for Suen's inequality.

Definition 3 (Dependency Graphs [7])**.** *Let* $\{V_i\}_{i\in\mathcal{I}}$ *be a family of random variables (defined on a common probability space). A dependency graph for* $\{V_i\}$ *is any graph* Γ *with vertex set* $V(\Gamma) = \mathcal{I}$ *whose set of edges satisfies the following property: if* A *and* B *are two disjoint subsets of* $\mathcal{I}$ *such that there are no edges with one vertex in* A *and the other in* B*, then the families* $\{V_i\}_{i\in A}$ *and* $\{V_i\}_{i\in B}$ *are independent.*

Fact 2. *[7, Example* 1.5*, p.11] Suppose* $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$ *is a family of independent random variables, and each* $V_i, i \in \mathcal{I}$ *is a function of the variables* $\{Y_\alpha\}_{\alpha\in A_i}$ *for some subset* $A_i \subseteq \mathcal{A}$*. Then the graph with vertex set* $\mathcal{I}$ *and edge set* $\{ij : A_i \cap A_j \neq \emptyset\}$ *is a dependency graph for* $\{U_i\}_{i\in\mathcal{I}}$*.*

In our setting, we fix $\epsilon = L^{-3/2}$, let V_i be the indicator the random variable defined in (57). Note that V_i is one if and only if $\beta(i)$ is an ϵ-good solution. The set of codewords that share at least one common term with $\beta(i)$ are the ones that play a role in determining whether $\beta(i)$ is an ϵ-good solution or not. Hence, the graph Γ with vertex set $V(\Gamma) = \{1, \ldots, e^{nR}\}$ and edge set $e(\Gamma)$ given by

$$\begin{aligned}\{ij : i \neq j \text{ and the codewords } \beta(i), \beta(j)\\ \text{share at least one common term}\}\end{aligned}$$

is a dependency graph for the family $\{V_i\}_{i=1}^{e^{nR}}$. This follows from Fact 2 by observing that: i) each V_i is a function of the

columns of $\mathbf{A}$ that define $\beta(i)$ and all other codewords that share at least one common term with $\beta(i)$; and ii) the columns of $\mathbf{A}$ are generated independently of one another.

For a given codeword $\beta(i)$, there are $\binom{L}{r}(M-1)^{L-r}$ other codewords that have exactly r terms in common with $\beta(i)$, for $0 \leq r \leq (L-1)$. Therefore each vertex in the dependency graph for the family $\{V_i\}_{i=1}^{e^{nR}}$ is connected to

$$\sum_{r=1}^{L-1} \binom{L}{r}(M-1)^{L-r} = M^L - 1 - (M-1)^L$$

other vertices.

Fact 3 (Suen's Inequality [7])**.** *Let $V_i \sim$ Bern$(p_i), i \in \mathcal{I}$, be a finite family of Bernoulli random variables having a dependency graph Γ. Write $i \sim j$ if ij is an edge in Γ. Define*

$$\lambda = \sum_{i\in\mathcal{I}} \mathbb{E}V_i, \;\; \Delta = \frac{1}{2}\sum_{i\in\mathcal{I}}\sum_{j\sim i} \mathbb{E}(V_iV_j), \;\; \delta = \max_{i\in\mathcal{I}} \sum_{k\sim i} \mathbb{E}V_k.$$

Then

$$P\left(\sum_{i\in\mathcal{I}} V_i = 0\right) \leq \exp\left(-\min\left\{\frac{\lambda}{2}, \frac{\lambda}{6\delta}, \frac{\lambda^2}{8\Delta}\right\}\right). \quad (69)$$

We apply Suen's inequality with the dependency graph specified above for $\{V_i\}_{i=1}^{e^{nR}}$ to compute an upper bound for $P(X_g = 0)$, where $X_g = \sum_{i=1}^{e^{nR}} V_i$ is the total number of ϵ-good solutions for $\epsilon = L^{-3/2}$. Note that the ϵ chosen here is smaller than the value of $L^{-1/2}$ used for Theorem 1. This smaller value is required to prove the super-exponential decay of the excess-distortion probability via Suen's inequality. We also need a stronger version of Lemma 2.

Lemma 4. *Let $R > \frac{1}{2}\log\frac{\rho^2}{D}$. If $\beta \in \mathcal{B}_{M,L}$ is a solution, then for sufficiently large L*

$$P\left(X_\alpha(\beta) \leq L^{-5/2}\,\mathbb{E}X, \;\; \textit{for } \tfrac{1}{L} \leq \alpha \leq \tfrac{L}{L}\right) \geq 1 - \xi \quad (70)$$

where

$$\xi = L^{-2.5\left(\frac{b}{b_{min}(\rho^2/D)} - \frac{7}{5}\right)}. \quad (71)$$

Proof: The proof is nearly identical to that of Lemma 2 given in Section V, with the terms $L^{-3/2}$ and $\frac{3}{2L}$ replaced by $L^{-5/2}$ and $\frac{5}{2L}$, respectively, throughout the lemma. Thus we obtain the following condition on b which is the analog of (107).

$$\begin{aligned} b > & \max_{\alpha\in\{\frac{1}{L},\ldots,\frac{L}{L}\}} \left\{\frac{R}{(\min\{\alpha\Lambda(\alpha), c_1\})}\left[\min\left\{\alpha, \bar{\alpha}, \frac{\log 2}{\log L}\right\} + \frac{5}{2L}\right]\right\} \\ &= \frac{3.5R}{\Lambda(0)} + O\left(\frac{1}{L}\right) \\ &= \frac{7}{5} b_{min}\left(\frac{\rho^2}{D}\right) + O\left(\frac{1}{L}\right). \end{aligned} \quad (72)$$

The result is then obtained using arguments analogous to (108) and (109). ∎

We now compute each of the three terms in the RHS of Suen's inequality.

First Term $\frac{\lambda}{2}$: We have

$$\begin{aligned} \lambda &= \sum_{i=1}^{e^{nR}} \mathbb{E}V_i = \mathbb{E}X_g \\ &\overset{(a)}{=} \mathbb{E}X \cdot P(\beta(1) \text{ is } \epsilon\text{-good } \mid \beta(1) \text{ is a solution}), \end{aligned} \quad (73)$$

where (a) follows from (60). Given that $\beta(1)$ is a solution, Lemma 4 shows that

$$\sum_{\alpha=\frac{1}{L},\ldots,\frac{L}{L}} X_\alpha(\beta(1)) < (\mathbb{E}X)L^{-3/2} \quad (74)$$

with probability at least $1-\xi$. As $\epsilon = L^{-3/2}$, $\beta(1)$ is ϵ-good according to Definition 2 if (74) is satisfied. Thus the RHS of (73) can be lower bounded as follows.

$$\begin{aligned} \lambda &= \mathbb{E}X \cdot P(\beta(1) \text{ is } \epsilon\text{-good } \mid \beta(1) \text{ is a solution}) \\ &\geq \mathbb{E}X \cdot (1-\xi). \end{aligned} \quad (75)$$

Using the expression from (33) for the expected number of solutions $\mathbb{E}X$, we have

$$\lambda \geq (1-\xi)\frac{\kappa}{\sqrt{n}} e^{n(R - \frac{1}{2}\log\frac{\rho^2}{D})}, \quad (76)$$

where $\kappa > 0$ is a constant. For $b > \frac{7}{5} b_{min}(\rho^2/D)$, (71) implies that ξ approaches 1 with growing L.

Second term $\lambda/(6\delta)$: Due to the symmetry of the code construction, we have

$$\begin{aligned} \delta &= \max_{i\in\{1,\ldots,e^{nR}\}} \sum_{k\sim i} P\left(V_k = 1\right) \\ &= \sum_{k\sim i} P\left(V_k = 1\right) \quad \forall i \in \{1,\ldots,e^{nR}\} \\ &= \sum_{r=1}^{L-1} \binom{L}{r}(M-1)^{L-r} \cdot P\left(V_1 = 1\right) \\ &= \left(M^L - 1 - (M-1)^L\right) P\left(V_1 = 1\right). \end{aligned} \quad (77)$$

Combining this together with the fact that

$$\lambda = \sum_{i=1}^{M^L} \mathbb{E}V_i = M^L\, P(V_1 = 1),$$

we obtain

$$\frac{\lambda}{\delta} = \frac{M^L}{M^L - 1 - (M-1)^L} = \frac{1}{1 - L^{-bL} - (1-L^{-b})^L}, \quad (78)$$

where the second equality is obtained by substituting $M = L^b$. Using a Taylor series bound for the denominator of (78) (see [3, Sec. V] for details) yields the following lower bound for sufficiently large L:

$$\frac{\lambda}{\delta} \geq \frac{L^{b-1}}{2}. \quad (79)$$

Third Term $\lambda^2/(8\Delta)$: We have

$$\begin{aligned}\Delta &= \frac{1}{2}\sum_{i=1}^{M^L}\sum_{j\sim i}\mathbb{E}\left[V_iV_j\right]\\ &= \frac{1}{2}\sum_{i=1}^{M^L}P(V_i=1)\sum_{j\sim i}P(V_j=1\mid V_i=1)\\ &\overset{(a)}{=} \frac{1}{2}\,\mathbb{E}X_g\sum_{j\sim 1}P(V_j=1\mid V_1=1)\\ &= \frac{1}{2}\,\mathbb{E}X_g\,\mathbb{E}\Big[\sum_{j\sim 1}\mathbf{1}\{V_j=1\}\mid V_1=1\Big]\\ &\overset{(b)}{\leq} \frac{1}{2}\,\mathbb{E}X_g\,\mathbb{E}\left[\sum_{\alpha=\frac{1}{L},\ldots,\frac{L-1}{L}}X_\alpha(\beta(1))\mid V_1=1\right].\end{aligned} \tag{80}$$

In (80), (a) holds because of the symmetry of the code construction. The inequality (b) is obtained as follows. The number of ϵ-good solutions that share common terms with $\beta(1)$ is bounded above by the total number of solutions sharing common terms with $\beta(1)$. The latter quantity can be expressed as the sum of the number of solutions sharing exactly αL common terms with $\beta(1)$, for $\alpha \in \{\frac{1}{L},\ldots,\frac{L-1}{L}\}$.

Conditioned on $V_1=1$, i.e., the event that $\beta(1)$ is a ϵ-good solution, the total number of solutions that share common terms with $\beta(1)$ is bounded by $\epsilon\,\mathbb{E}X$. Therefore, from (80) we have

$$\begin{aligned}\Delta &\leq \frac{1}{2}\mathbb{E}X_g\,\mathbb{E}\left[\sum_{\alpha=\frac{1}{L},\ldots,\frac{L-1}{L}}X_\alpha(\beta(1))\mid V_1=1\right]\\ &\leq \frac{1}{2}\left(\mathbb{E}X_g\right)\left(L^{-3/2}\,\mathbb{E}X\right) \leq \frac{L^{-3/2}}{2}(\mathbb{E}X)^2,\end{aligned} \tag{81}$$

where we have used $\epsilon = L^{-3/2}$, and the fact that $X_g \leq X$. Combining (81) and (75), we obtain

$$\frac{\lambda^2}{8\Delta} \geq \frac{(1-\xi)^2(\mathbb{E}X)^2}{4L^{-3/2}(\mathbb{E}X)^2} \geq \kappa L^{3/2}, \tag{82}$$

where κ is a strictly positive constant.

Applying Suen's inequality: Using the lower bounds obtained in (76), (79), and (82) in (69), we obtain

$$\begin{aligned}&P\left(\sum_{i=1}^{e^{nR}}V_i\right)\\ &\leq \exp\left(-\kappa\,\min\left\{e^{n(R-\frac{1}{2}\log\frac{\rho^2}{D}-\frac{\log n}{2n})},\,L^{b-1},\,L^{3/2}\right\}\right),\end{aligned} \tag{83}$$

where κ is a positive constant. Recalling from (3) that $L = \Theta(\frac{n}{\log n})$ and $R > \frac{1}{2}\ln\frac{\rho^2}{D}$, we see that for $b > 2$,

$$P\left(\sum_{i=1}^{e^{nR}}V_i\right) \leq \exp\left(-\kappa n^{1+c}\right), \tag{84}$$

where $c > 0$ is a constant. Note that the condition $b > \frac{7}{5}b_{min}(\rho^2/D)$ was also needed to obtain (83) via Suen's inequality. In particular, this condition on b is required for ξ in Lemma 4 to go to 0 with growing L.

Using (84) in (19), we conclude that for any $\gamma^2 \in (\sigma^2, D^{e^{2R}})$ the probability of excess distortion can be bounded as

$$\begin{aligned}P_{e,n} &\leq P(|\mathbf{S}|^2 \geq \gamma^2) + \max_{\rho^2\in(D,\gamma^2)}P(\mathcal{E}(\tilde{\mathbf{S}})\mid|\tilde{\mathbf{S}}|^2=\rho^2)\\ &\leq P(|\mathbf{S}|^2\geq\gamma^2)+\exp(-\kappa n^{1+c}),\end{aligned} \tag{85}$$

provided the parameter b satisfies

$$b > \max_{\rho^2\in(D,\gamma^2)}\max\left\{2,\ \frac{7}{5}b_{min}\left(\rho^2/D\right)\right\}. \tag{86}$$

It can be verified from the definition in (7) that $b_{min}(x)$ is strictly increasing in $x \in (1, e^{2R})$. Therefore, the maximum on the RHS of (86) is bounded by $\max\left\{2,\ \frac{7}{5}b_{min}\left(\gamma^2/D\right)\right\}$. Choosing b to be larger than this value will guarantee that (85) holds. This completes the proof of the theorem.

V. Proof of Lemma 2

We begin by listing three useful properties of the function $f(x,y,z)$ defined in (27). Recall that the probability that an i.i.d. $\mathcal{N}(0,y)$ sequence is within distortion within distortion z of a norm-x sequence is $\sim e^{-nf(x,y,z)}$.

1) For fixed x, y, f is strictly decreasing in $z \in (0, x+y)$.
2) For fixed y, z, f is strictly increasing in $x \in (z, \infty)$.
3) For fixed x, z and $x > z$, f is convex in y and attains its minimum value of $\frac{1}{2}\log\frac{x}{z}$ at $y = x - z$.

These properties are straightforward to verify from the definition (27) using elementary calculus.

For $\mathcal{K} \subseteq \{1,\ldots,L\}$, let $\beta_{\mathcal{K}}$ denote the restriction of β to the set $\mathcal{K}$, i.e., $\beta_{\mathcal{K}}$ coincides with β in the sections indicated by $\mathcal{K}$ and the remaining entries are all equal to zero. For example, if $\mathcal{K} = \{2,3\}$, the second and third sections of $\beta_{\mathcal{K}}$ will each have one non-zero entry, the other entries are all zeros.

Definition 4. *Given that β is a solution, for $\alpha = \frac{1}{L},\ldots,\frac{L}{L}$, define $\mathcal{F}_\alpha(\beta)$ as the event that*

$$|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 \geq D_\alpha$$

for every size αL subset $\mathcal{K} \subset \{1,\ldots,L\}$, where D_α is the solution to the equation

$$R\alpha = f(\rho^2, (\rho^2 - D)\alpha, D_\alpha). \tag{87}$$

The intuition behind choosing D_α according to (87) is the following. Any subset of αL sections of the design matrix $\mathbf{A}$ defines a SPARC of rate $R\alpha$, with each codeword consisting of i.i.d $\mathcal{N}(0,(\rho^2-D)\alpha)$ entries. (Note that the entries of a single codeword are i.i.d., though the codewords are dependent due to the SPARC structure.) The probability that a codeword from this rate $R\alpha$ code is within distortion z of the source sequence $\tilde{\mathbf{S}}$ is $\sim e^{-nf(\rho^2,(\rho^2-D)\alpha,z)}$. Hence the expected number of codewords in the rate $R\alpha$ codebook within distortion z of $\tilde{\mathbf{S}}$ is

$$e^{nR\alpha}e^{-nf(\rho^2,(\rho^2-D)\alpha,z)}.$$

As $f(\rho^2,(\rho^2-D)\alpha,z)$ is a strictly decreasing function of z in $(0,\rho^2)$, (87) says that D_α is the smallest expected distortion

for any rate $R\alpha$ code with codeword entries chosen i.i.d. $\mathcal{N}(0,(\rho^2-D)\alpha)$. [4] For $z < D_\alpha$, the expected number of codewords within distortion z of $\tilde{\mathbf{S}}$ is vanishingly small.

Conditioned on $\mathcal{F}_\alpha(\beta)$, the idea is that any αL sections of β cannot by themselves represent $\tilde{\mathbf{S}}$ with distortion less than D_α. In other words, in a typical realization of the design matrix, all the sections contribute roughly equal amounts to finding a codeword within D of $\tilde{\mathbf{S}}$. On the other hand, if some αL sections of the SPARC can represent $\tilde{\mathbf{S}}$ with distortion less than D_α, the remaining $\bar{\alpha}L$ sections have "less work" to do—this creates a proliferation of solutions that share these αL common sections with β. Consequently, the total number of solutions is much greater than $\mathbb{E}X$ for these atypical design matrices.

The first step in proving the lemma is to show that for any β, the event $\mathcal{F}_\alpha(\beta)$ holds w.h.p. The second step is showing that when $\mathcal{F}_\alpha(\beta)$ holds, the expected number of solutions that share any common terms with β is small compared to $\mathbb{E}X$. Indeed, using $\mathcal{F}_\alpha(\beta)$ we can write

$$\begin{aligned}
&P\left(X_\alpha(\beta) > L^{-3/2}\mathbb{E}X\right)\\
&= P\left(\{X_\alpha(\beta) > L^{-3/2}\mathbb{E}X\},\ \mathcal{F}_\alpha^c(\beta)\right)\\
&\quad + P\left(\{X_\alpha(\beta) > L^{-3/2}\,\mathbb{E}X\},\ \mathcal{F}_\alpha(\beta)\right)\\
&\le P(\mathcal{F}_\alpha^c(\beta)) + P(\mathcal{F}_\alpha(\beta))P\left(X_\alpha(\beta) < L^{-3/2}\mathbb{E}X \mid \mathcal{F}_\alpha(\beta)\right)\\
&\le P(\mathcal{F}_\alpha^c(\beta)) + \frac{\mathbb{E}[X_\alpha(\beta)\mid \mathcal{F}_\alpha(\beta)]}{L^{-3/2}\,\mathbb{E}X}
\end{aligned} \tag{88}$$

where the last line follows from Markov's inequality. We will show that the probability on the left side of (88) is small for any solution β by showing that each of the two terms on the RHS of (88) is small. First, a bound on D_α.

Lemma 5. *For $\alpha \in (0,1]$,*

$$R\alpha > f(\rho^2,\,(\rho^2-D)\alpha,\,\rho^2\bar{\alpha}+D\alpha) = \frac{1}{2}\log\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}. \tag{89}$$

Consequently, $D_\alpha < \rho^2\bar{\alpha} + D\alpha$ for $\alpha = \frac{1}{L},\ldots,\frac{L}{L}$.

Proof. The last equality in (89) holds because $f(x,x-z,z) = \frac{1}{2}\ln\frac{x}{z}$. Define a function

$$g(\alpha) = R\alpha - \frac{1}{2}\log\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}.$$

Then $g(0) = 0$, $g(1) = R - \frac{1}{2}\ln\frac{\rho^2}{D} > 0$, and the second derivative is

$$\frac{d^2g}{d\alpha^2} = \frac{-(1-\frac{D}{\rho^2})^2}{(1-(1-\frac{D}{\rho^2})\alpha)^2} < 0.$$

Therefore g is strictly concave in $[0,1]$, and its minimum value (at $\alpha = 0$) is 0. This proves (89). Recalling the definition of D_α in (87), (89) implies that

$$f(\rho^2,(\rho^2-D)\alpha,D_\alpha) = R\alpha > f(\rho^2,\,(\rho^2-D)\alpha,\,\rho^2\bar{\alpha}+D\alpha)$$

As f is decreasing in its third argument (the distortion), we conclude that $D_\alpha < \rho^2\bar{\alpha} + D\alpha$. □

We now bound each term on the RHS of (88). Showing that the first term of (88) is small implies that w.h.p any αL sections by themselves will leave a residual distortion of at least D_α. Showing that the second term is small implies that under this condition, the expected number of solutions sharing any common terms with β is small compared to $\mathbb{E}X$.

Bounding $\mathcal{F}_\alpha^c(\beta)$: From the definition of the event $\mathcal{F}_\alpha(\beta)$, we have

$$P(\mathcal{F}_\alpha^c(\beta)) = \cup_{\mathcal{K}}\ P(|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 < D_\alpha \mid \beta \text{ is a solution}) \tag{90}$$

where the union is over all size-αL subsets of $\{1,\ldots,L\}$. Using a union bound, (90) becomes

$$P(\mathcal{F}_\alpha^c(\beta)) \le \binom{L}{L\alpha}\frac{P(|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 < D_\alpha,\,|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D)}{P(|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D)} \tag{91}$$

where $\mathcal{K}$ is a generic size-αL subset of $\{1,\ldots,L\}$, say $\mathcal{K} = \{1,\ldots,\alpha L\}$. Recall from (33) that for sufficiently large n, the denominator in (91) can be bounded from below as

$$P(|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D) \ge \frac{\kappa}{\sqrt{n}}e^{-nf(\rho^2,\rho^2-D,D)} \tag{92}$$

and $f(\rho^2,\rho^2-D,D) = \frac{1}{2}\log\frac{\rho^2}{D}$. The numerator in (91) can be expressed as

$$\begin{aligned}
&P(|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 < D_\alpha,\,|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D)\\
&= \int_0^{D_\alpha}\psi(y)\,P(|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D \mid |\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 = y)\,dy
\end{aligned} \tag{93}$$

where ψ is the density of the random variable $|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2$. Using the cdf at y to bound $\psi(y)$ in the RHS of (93), we obtain the following upper bound for sufficiently large n.

$$\begin{aligned}
&P(|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 < D_\alpha,\,|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D)\\
&\le \int_0^{D_\alpha} P(|\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 < y)\\
&\qquad\qquad \cdot P(|\tilde{\mathbf{S}} - \mathbf{A}\beta|^2 < D \mid |\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 = y)\,dy\\
&\overset{(a)}{\le} \int_0^{D_\alpha}\frac{\kappa}{\sqrt{n}}e^{-nf(\rho^2,(\rho^2-D)\alpha,y)}\\
&\qquad \cdot P(|(\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}) - \mathbf{A}\beta_{\mathcal{K}^c}|^2 < D \mid |\tilde{\mathbf{S}} - \mathbf{A}\beta_{\mathcal{K}}|^2 = y)\,dy\\
&\overset{(b)}{\le} \int_0^{D_\alpha}\frac{\kappa}{\sqrt{n}}e^{-nf(\rho^2,(\rho^2-D)\alpha,y)}\cdot e^{-nf(y,(\rho^2-D)\bar{\alpha},D)}\,dy\\
&\overset{(c)}{\le} \int_0^{D_\alpha}\frac{\kappa}{\sqrt{n}}e^{-nf(\rho^2,(\rho^2-D)\alpha,D_\alpha)}\cdot e^{-nf(D_\alpha,(\rho^2-D)\bar{\alpha},D)}\,dy.
\end{aligned} \tag{94}$$

In (94), (a) holds for sufficiently large n and is obtained using the strong version of Cramér's large deviation theorem: note that $\mathbf{A}\beta_{\mathcal{K}}$ is a linear combination of αL columns of $\mathbf{A}$, hence it is a Gaussian random vector with i.i.d. $\mathcal{N}(0,(\rho^2-D)\alpha)$ entries that is independent of $\tilde{\mathbf{S}}$. Inequality (b) is similarly obtained: $\mathbf{A}\beta_{\mathcal{K}^c}$ has i.i.d. $\mathcal{N}(0,(\rho^2-D)\bar{\alpha})$ entries, and is independent of both $\tilde{\mathbf{S}}$ and $\mathbf{A}\beta_{\mathcal{K}}$. Finally, (c) holds because

[4] Note that D_α is *not* the distortion-rate function at rate $R\alpha$ as the codewords are not chosen with the optimal variance for rate $R\alpha$.

the overall exponent

$$f(\rho^2,(\rho^2-D)\alpha,y)+f(y,(\rho^2-D)\bar{\alpha},D)$$

is a decreasing function of y, for $y \in (0, \rho^2\bar{\alpha}+D\alpha]$, and $D_\alpha \le \rho^2\bar{\alpha}+D\alpha$.

Using (92) and (94) in (91), for sufficiently large n we have

$$P(\mathcal{F}_\alpha^c(\beta)) \le \kappa\binom{L}{L\alpha} \times e^{-n[f(\rho^2,(\rho^2-D)\alpha,D_\alpha)+f(D_\alpha,(\rho^2-D)\bar{\alpha},D)-f(\rho^2,\rho^2-D,D)]}. \tag{95}$$

Bounding $\mathbb{E}[X_\alpha(\beta)|\,\mathcal{F}_\alpha(\beta)]$: There are $\binom{L}{L\alpha}(M-1)^{L\bar{\alpha}}$ codewords which share αL common terms with β. Therefore

$$\mathbb{E}[X_\alpha(\beta)\mid\mathcal{F}_\alpha(\beta)] = \binom{L}{L\alpha}(M-1)^{L\bar{\alpha}} \times P(|\tilde{\mathbf{S}}-\mathbf{A}\beta'|^2<D \mid |\tilde{\mathbf{S}}-\mathbf{A}\beta|^2<D,\,\mathcal{F}_\alpha(\beta)) \tag{96}$$

where β' is a codeword that shares exactly αL common terms with β. If $\mathcal{K}$ is the size-αL set of common sections between β and β', then $\beta' = \beta_{\mathcal{K}} + \beta'_{\mathcal{K}^c}$ and

$$\begin{aligned}
&P(|\tilde{\mathbf{S}}-\mathbf{A}\beta'|^2<D \mid |\tilde{\mathbf{S}}-\mathbf{A}\beta|^2<D,\,\mathcal{F}_\alpha(\beta))\\
&= P(|(\tilde{\mathbf{S}}-\mathbf{A}\beta_{\mathcal{K}})-\mathbf{A}\beta'_{\mathcal{K}^c}|^2<D \mid |\tilde{\mathbf{S}}-\mathbf{A}\beta|^2<D,\,\mathcal{F}_\alpha(\beta))\\
&\overset{(a)}{\le} P\left(\tfrac{1}{n}\sum_{i=1}^{n}(D_\alpha-(\mathbf{A}\beta'_{\mathcal{K}^c})_i)^2 \;<D\right)\\
&\overset{(b)}{\le} \frac{\kappa}{\sqrt{n}}e^{-nf(D_\alpha,(\rho^2-D)\bar{\alpha},D)},
\end{aligned} \tag{97}$$

where (b) holds for sufficiently large n. In (97), (a) is obtained as follows. Under the event $\mathcal{F}_\alpha(\beta)$, the norm $|\tilde{\mathbf{S}}-\mathbf{A}\beta_{\mathcal{K}}|^2$ is at least D_α, and $\mathbf{A}\beta'_{\mathcal{K}^c}$ is an i.i.d. $\mathcal{N}(0,(\rho^2-D)\bar{\alpha})$ vector independent of $\tilde{\mathbf{S}}$, β, and $\beta_{\mathcal{K}}$. (a) then follows from the rotational invariance of the distribution of $\mathbf{A}\beta'_{\mathcal{K}^c}$. Inequality (b) is obtained using the strong version of Cramér's large deviation theorem.

Using (97) in (96), we obtain for sufficiently large n

$$\begin{aligned}
&\mathbb{E}[X_\alpha(\beta)\mid\mathcal{F}_\alpha(\beta)]\\
&\le \binom{L}{L\alpha}(M-1)^{L\bar{\alpha}}\frac{\kappa}{\sqrt{n}}e^{-nf(D_\alpha,(\rho^2-D)\bar{\alpha},D)}\\
&\le \binom{L}{L\alpha}\frac{\kappa}{\sqrt{n}}e^{n(R\bar{\alpha}-f(D_\alpha,(\rho^2-D)\bar{\alpha},D))}.
\end{aligned} \tag{98}$$

Overall bound: Substituting the bounds from (95), (98) and (33) in (88), for sufficiently large n we have for $\frac{1}{L}\le\alpha\le 1$:

$$\begin{aligned}
&P\left(X_\alpha(\beta)>L^{-3/2}\mathbb{E}X\right) \le \kappa\binom{L}{L\alpha}\\
&\times\Big(e^{-n[f(\rho^2,(\rho^2-D)\alpha,D_\alpha)+f(D_\alpha,(\rho^2-D)\bar{\alpha},D)-f(\rho^2,\rho^2-D,D)]}\\
&+L^{3/2}e^{-n[R\alpha+f(D_\alpha,(\rho^2-D)\bar{\alpha},D)-f(\rho^2,(\rho^2-D),D)]}\Big).
\end{aligned} \tag{99}$$

Since D_α is chosen to satisfy $R\alpha = f(D_\alpha,(\rho^2-D)\bar{\alpha},D)$, the two exponents in (99) are equal. To bound (99), we use the following lemma.

Lemma 6. *For $\alpha\in\{\frac{1}{L},\dots,\frac{L-1}{L}\}$, we have*

$$\begin{aligned}
&\Big[f(\rho^2,(\rho^2-D)\alpha,D_\alpha)+f(D_\alpha,(\rho^2-D)\bar{\alpha},D)\\
&\quad-f(\rho^2,(\rho^2-D),D)\Big]\\
&\quad>\begin{cases}\alpha\Lambda(\alpha) & \text{if } D_\alpha>D\\ c_1 & \text{if } D_\alpha\le D.\end{cases}
\end{aligned} \tag{100}$$

where D_α is the solution of (87)*, c_1 is a positive constant given by* (137)*, and*

$$\begin{aligned}
\Lambda(\alpha) &= \frac{1}{8}\left(\frac{D}{\rho^2}\right)^4\left(1+\frac{D}{\rho^2}\right)^2\left(1-\frac{D}{\rho^2}\right)\\
&\cdot\left[-1+\left(1+\frac{2\sqrt{\rho^2/D}}{(\frac{\rho^2}{D}-1)}\left(R-\frac{1}{2\alpha}\log\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}\right)\right)^{\frac{1}{2}}\right]^2.
\end{aligned} \tag{101}$$

Proof: See Appendix I. ∎

We observe that $\Lambda(\alpha)$ is strictly decreasing for $\alpha\in(0,1]$. This can be seen by using the Taylor expansion of $\log(1-x)$ for $0<x<1$ to write

$$\begin{aligned}
R-\frac{1}{2\alpha}\ln\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha} &= R+\frac{1}{2\alpha}\log\left(1-\alpha\left(1-\frac{D}{\rho^2}\right)\right)\\
&= R-\frac{1}{2}\sum_{k=1}^{\infty}\left(1-\frac{D}{\rho^2}\right)^k\frac{\alpha^{k-1}}{k}.
\end{aligned} \tag{102}$$

Since

$$R>\frac{1}{2}\log\frac{\rho^2}{D}>\frac{1}{2}\left(1-\frac{D}{\rho^2}\right),$$

(102) shows that $\Lambda(\alpha)$ is strictly positive and strictly decreasing in $\alpha\in(0,1)$ with

$$\begin{aligned}
\Lambda(0) &:= \lim_{\alpha\to0}\Lambda(\alpha) = \frac{1}{8}\left(\frac{D}{\rho^2}\right)^4\left(1+\frac{D}{\rho^2}\right)^2\left(1-\frac{D}{\rho^2}\right)\\
&\left[-1+\left(1+\frac{2\sqrt{\rho^2/D}}{(\frac{\rho^2}{D}-1)}\left(R-\frac{1}{2}\left(1-\frac{D}{\rho^2}\right)\right)\right)^{\frac{1}{2}}\right]^2,\\
\Lambda(1) &= \frac{1}{8}\left(\frac{D}{\rho^2}\right)^4\left(1+\frac{D}{\rho^2}\right)^2\left(1-\frac{D}{\rho^2}\right)\\
&\left[-1+\left(1+\frac{2\sqrt{\rho^2/D}}{(\frac{\rho^2}{D}-1)}\left(R-\frac{1}{2}\log\frac{\rho^2}{D}\right)\right)^{\frac{1}{2}}\right]^2.
\end{aligned} \tag{103}$$

Substituting (100) in (99), we have, for $\alpha\in\{\frac{1}{L},\dots,\frac{L}{L}\}$:

$$\begin{aligned}
&P\left(X_\alpha(\beta)>L^{-3/2}\mathbb{E}X\right)\\
&<\kappa\binom{L}{L\alpha}L^{3/2}\exp(-n\cdot\min\{\alpha\Lambda(\alpha),c_1\}).
\end{aligned} \tag{104}$$

Taking logarithms and dividing both sides by $L\log L$, we obtain

$$\begin{aligned}&\frac{1}{L\log L}\log P\left(X_\alpha(\beta)>L^{-3/2}\mathbb{E}X\right)\\&<\frac{\log\kappa}{L\log L}+\frac{\log\binom{L}{L\alpha}}{L\log L}+\frac{3}{2L}-\frac{n(\min\{\alpha\Lambda(\alpha),c_1\})}{L\log L}\\&\overset{(a)}{=}\frac{\log\kappa}{L\log L}+\min\left\{\alpha,\bar{\alpha},\frac{\log 2}{\log L}\right\}+\frac{3}{2L}\\&\quad-\frac{(\min\{\alpha\Lambda(\alpha),c_1\})b}{R}\end{aligned}\tag{105}$$

where to obtain (a), we have used the bound

$$\log\binom{L}{L\alpha}<\min\left\{\alpha L\log L,\ (1-\alpha)L\log L,\ L\log 2\right\}$$

and the relation (3). For the right side of (105) to be negative for sufficiently large L, we need

$$\frac{(\min\{\alpha\Lambda(\alpha),c_1\})b}{R}>\min\left\{\alpha,\bar{\alpha},\frac{\log 2}{\log L}\right\}+\frac{3}{2L}.\tag{106}$$

This can be arranged by choosing b to be large enough. Since (106) has to be satisfied for all $\alpha\in\{\frac{1}{L},\ldots,\frac{L-1}{L}\}$, we need

$$\begin{aligned}&b>\\&\max_{\alpha\in\{\frac{1}{L},\ldots,\frac{L}{L}\}}\left\{\frac{R}{(\min\{\alpha\Lambda(\alpha),c_1\})}\left[\min\left\{\alpha,\bar{\alpha},\frac{\log 2}{\log L}\right\}+\frac{3}{2L}\right]\right\}\\&\overset{(a)}{=}\frac{2.5R}{\Lambda(0)}+O\left(\frac{1}{L}\right)\\&=b_{min}\left(\frac{\rho^2}{D}\right)+O\left(\frac{1}{L}\right).\end{aligned}\tag{107}$$

In (107), (a) holds because $\Lambda(\alpha)$ is of constant order for all $\alpha\in(0,1]$, hence the maximum is attained at $\alpha=\frac{1}{L}$. The constant $\Lambda(0)$ is given by (103), and $b_{min}(.)$ is defined in the statement of Theorem 1.

When b satisfies (107) and L is sufficiently large, for $\alpha\in\{\frac{1}{L},\ldots,\frac{L}{L}\}$, the bound in (105) becomes

$$\begin{aligned}&\frac{1}{L\log L}\log P\left(X_\alpha(\beta)>L^{-3/2}\mathbb{E}X\right)\\&<\frac{\log\kappa}{L\log L}-\frac{\min\{\alpha\Lambda(\alpha),c_1\}(b-b_{min}-O(\frac{1}{L}))}{R}\\&\leq\frac{\log\kappa}{L\log L}-\frac{\Lambda(0)}{L}\frac{(b-b_{min})}{R}=\frac{\log\kappa}{L\log L}-\frac{2.5(\frac{b}{b_{min}}-1)}{L}.\end{aligned}\tag{108}$$

Therefore

$$P\left(X_\alpha(\beta)>L^{-3/2}\mathbb{E}X\right)<\kappa L^{-2.5(\frac{b}{b_{min}}-1)}.\tag{109}$$

This completes the proof of Lemma 2.

Appendix I
Proof of Lemma 6

For $\alpha\in\{\frac{1}{L},\ldots,\frac{L-1}{L}\}$, define the function $g_\alpha:\mathbb{R}\to\mathbb{R}$ as

$$g_\alpha(u)=f(\rho^2,(\rho^2-D)\alpha,u)+f(u,(\rho^2-D)\bar{\alpha},D)-\frac{1}{2}\ln\frac{\rho^2}{D}.\tag{110}$$

We want a lower bound for $g_\alpha(D_\alpha)\geq\Lambda(\alpha)\alpha$, where D_α is the solution to

$$R\alpha=f(\rho^2,(\rho^2-D)\alpha,D_\alpha).\tag{111}$$

We consider the cases $D_\alpha>D$ and $D_\alpha\leq D$ separately. Recall from Lemma 5 that $D_\alpha<\rho^2\bar{\alpha}+D\alpha$.

Case 1: $D<D_\alpha<\rho^2\bar{\alpha}+D\alpha$.: In this case, both the $f(.)$ terms in the definition of $g_\alpha(D_\alpha)$ are strictly positive. We can write

$$D_\alpha=\rho^2\bar{\alpha}+D\alpha-\delta,\tag{112}$$

where $\delta\in(0,(\rho^2-D)\bar{\alpha})$. Expanding $g_\alpha(\rho^2\bar{\alpha}+D\alpha-\delta)$ around $\rho^2\bar{\alpha}+D\alpha$ using Taylor's theorem, we obtain

$$g(D_\alpha)=g(\rho^2\bar{\alpha}+D\alpha)-g'(\rho^2\bar{\alpha}+D\alpha)\delta+g''(\xi)\frac{\delta^2}{2}=g''(\xi)\frac{\delta^2}{2}\tag{113}$$

since $g(\rho^2\bar{\alpha}+D\alpha)=g'(\rho^2\bar{\alpha}+D\alpha)=0$. Here ξ is a number in the interval $(D,\rho^2\bar{\alpha}+D\alpha)$. We bound $g(D_\alpha)$ from below by obtaining separate lower bounds for $g''(\xi)$ and δ.

Lower Bound for $g''(\xi)$: Using the definition of f in (27), the second derivative of $g(u)$ is

$$\begin{aligned}g''(u)=&\frac{-1}{2u^2}\\&+\frac{2\rho^4\cdot[(\rho^2-D)^2\alpha^2+4\rho^2u]^{-1/2}}{\alpha(\rho^2-D)\left[\sqrt{(\rho^2-D)^2\alpha^2+4\rho^2u}-(\rho^2-D)\alpha\right]^2}\\&+\frac{2D^2\cdot[(\rho^2-D)^2\bar{\alpha}^2+4Du]^{-1/2}}{\bar{\alpha}(\rho^2-D)\left[\sqrt{(\rho^2-D)^2\bar{\alpha}^2+4Du}-(\rho^2-D)\bar{\alpha}\right]^2}.\end{aligned}\tag{114}$$

It can be verified that $g''(u)$ is a decreasing function, and hence for $\xi\in(D,\rho^2\bar{\alpha}+D\alpha)$,

$$\begin{aligned}g''(\xi)&\geq g''(\rho^2\bar{\alpha}+D\alpha)\\&=\frac{-1}{2(\rho^2\bar{\alpha}+D\alpha)^2}\\&\quad+\frac{\rho^4}{2\alpha(\rho^2-D)(\rho^2\bar{\alpha}+D\alpha)^2(\rho^2(1+\bar{\alpha})+D\alpha)}\\&\quad+\frac{1}{2\bar{\alpha}(\rho^2-D)(\rho^2\bar{\alpha}+D(1+\alpha))}\\&=\frac{(\rho^2+D)}{2\alpha\bar{\alpha}(\rho^2-D)(\rho^2(1+\bar{\alpha})+D\alpha)(\rho^2\bar{\alpha}+D(1+\alpha))}\\&\geq\frac{1}{4\alpha\bar{\alpha}(\rho^2-D)\rho^2}.\end{aligned}\tag{115}$$

Lower bound for δ: From (111) and (112), note that δ is the solution to

$$R\alpha=f(\rho^2,(\rho^2-D)\alpha,\rho^2\bar{\alpha}+D\alpha-\delta).\tag{116}$$

Using Taylor's theorem for f in its third argument around the point $p:=(\rho^2,(\rho^2-D)\alpha,\rho^2\bar{\alpha}+D\alpha)$, we have

$$\begin{aligned}R\alpha&=f(p)-\frac{\partial f}{\partial z}\Big|_p\delta+\frac{\partial^2 f}{\partial z^2}\Big|_{\tilde{p}}\frac{\delta^2}{2}\\&=\frac{1}{2}\ln\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}+\frac{1}{2(\rho^2\bar{\alpha}+D\alpha)}\delta+\frac{1}{2}\frac{\partial^2 f}{\partial z^2}\Big|_{\tilde{p}}\delta^2,\end{aligned}\tag{117}$$

where $\tilde{p} = p = (\rho^2, (\rho^2 - D)\alpha, \tilde{z})$ for some $\tilde{z} \in (D, \rho^2\bar{\alpha} + D\alpha)$. As (117) is a quadratic in δ with positive coefficients for the δ and δ^2 terms, replacing the δ^2 coefficient with an upper bound and solving the resulting quadratic will yield a *lower bound* for δ. Since the function

$$\frac{\partial^2 f}{\partial z^2}\Big|_{(x,y,z)} = \frac{2x^2}{y\sqrt{y^2+4xz}(\sqrt{y^2+4xz}-y)^2} \tag{118}$$

is decreasing in z, the δ^2 coefficient can be bounded as follows.

$$\frac{1}{2}\frac{\partial^2 f}{\partial z^2}\Big|_{\tilde{p}=(\rho^2,(\rho^2-D)\alpha,\tilde{z})} \leq a^* := \frac{1}{2}\frac{\partial^2 f}{\partial z^2}\Big|_{(\rho^2,(\rho^2-D)\alpha,D)}, \tag{119}$$

where a^* can be computed to be

$$a^* = \rho^4\Bigg(\alpha(\rho^2-D)\sqrt{(\rho^2-D)^2\alpha^2+4\rho^2 D} \left[\sqrt{(\rho^2-D)^2\alpha^2+4\rho^2 D}-(\rho^2-D)\alpha\right]^2\Bigg)^{-1}. \tag{120}$$

Therefore we can obtain a lower bound for δ, denoted by $\underline{\delta}$, by solving the equation

$$\underline{\delta}^2\, a^* + \underline{\delta}\,\frac{1}{2(\rho^2\bar{\alpha}+D\alpha)} - \left(R\alpha - \frac{1}{2}\ln\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}\right) = 0. \tag{121}$$

We thus obtain

$$\delta > \underline{\delta} = \frac{1}{4(\rho^2\bar{\alpha}+D\alpha)a^*}\Bigg[-1+\left(1+16(\rho^2\bar{\alpha}+D\alpha)^2 a^*\alpha\left(R-\frac{1}{2\alpha}\ln\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}\right)\right)^{1/2}\Bigg]. \tag{122}$$

We now show that $\underline{\delta}$ can be bounded from below by $\alpha\Lambda(\alpha)$ by obtaining lower and upper bounds for $a^*\alpha$. From (120) we have

$$a^*\alpha = \frac{\rho^4\cdot[(\rho^2-D)^2\alpha^2+4\rho^2 D]^{-1/2}}{(\rho^2-D)\left[\sqrt{(\rho^2-D)^2\alpha^2+4\rho^2 D}-(\rho^2-D)\alpha\right]^2} \geq \frac{\sqrt{\rho^2/D}}{8D(\rho^2-D)}, \tag{123}$$

where the inequality is obtained by noting that $a^*\alpha$ is strictly increasing in α, and hence taking $\alpha = 0$ gives a lower bound. Analogously, taking $\alpha = 1$ yields the upper bound

$$a^*\alpha \leq \frac{\rho^4}{4D^2(\rho^4-D^2)}. \tag{124}$$

Using the bounds of (123) and (124) in (122), we obtain

$$\delta > \underline{\delta} \geq \alpha\,\frac{D^2(\rho^4-D^2)}{\rho^6}\Bigg[-1+\left(1+\frac{2\sqrt{\rho^2/D}}{(\frac{\rho^2}{D}-1)}\left(R-\frac{1}{2\alpha}\ln\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}\right)\right)^{1/2}\Bigg]. \tag{125}$$

Finally, using the lower bounds for $g''(\xi)$ and δ from (125) and (115) in (113), we obtain

$$g(D_\alpha) > \frac{\alpha}{8}\left(\frac{D}{\rho^2}\right)^4\left(1+\frac{D}{\rho^2}\right)^2\left(1-\frac{D}{\rho^2}\right) \times\left[-1+\left(1+\frac{2\sqrt{\rho^2/D}}{(\frac{\rho^2}{D}-1)}\left(R-\frac{1}{2\alpha}\ln\frac{\rho^2}{\rho^2\bar{\alpha}+D\alpha}\right)\right)^{\frac{1}{2}}\right]^2 = \alpha\Lambda(\alpha). \tag{126}$$

Case 2: $D_\alpha \leq D$.: In this case, $g(D_\alpha)$ is given by

$$g(D_\alpha) = f(\rho^2,(\rho^2-D)\alpha,D_\alpha) + f(D_\alpha,(\rho^2-D)\bar{\alpha},D) - \frac{1}{2}\ln\frac{\rho^2}{D} = R\alpha - \frac{1}{2}\ln\frac{\rho^2}{D}, \tag{127}$$

where we have used (111) and the fact that $f(D_\alpha,(\rho^2 - D)\bar{\alpha},D) = 0$ for $D_\alpha \leq D$. The right hand side of the equation

$$R\alpha = f(\rho^2,(\rho^2-D)\alpha,z)$$

is decreasing in z for $z \in (0, D]$. Therefore, it is sufficient to consider $D_\alpha = D$ in order to obtain a lower bound for $R\alpha$ that holds for all $D_\alpha \in (0, D]$.

Next, we claim that the α that solves the equation

$$R\alpha = f(\rho^2,(\rho^2-D)\alpha,D) \tag{128}$$

lies in the interval $(\frac{1}{2},1)$. Indeed, observe that the LHS of (128) is increasing in α, while the RHS is decreasing in α for $\alpha \in (0,1]$. Since the LHS is strictly greater than the RHS at $\alpha = 1$ ($R > \frac{1}{2}\ln\frac{\rho^2}{D}$), the solution is strictly less than 1. On the other hand, for $\alpha \leq \frac{1}{2}$, we have

$$R\alpha \leq \frac{R}{2} \leq \frac{1}{2}\left(1-\frac{D}{\rho^2}\right) < \frac{1}{2}\ln\frac{\rho^2}{D} = f(\rho^2,(\rho^2-D),D) < f(\rho^2,\tfrac{(\rho^2-D)}{2},D), \tag{129}$$

i.e., the LHS of (128) is strictly less than the RHS. Therefore the α that solves (128) lies in $(\frac{1}{2},1)$.

To obtain a lower bound on the RHS of (127), we expand $f(\rho^2,(\rho^2 - D)\alpha,D)$ using Taylor's theorem for the second argument.

$$f(\rho^2,(\rho^2-D)\alpha,D) = f(\rho^2,(\rho^2-D)-\Delta,D) = \frac{1}{2}\ln\frac{\rho^2}{D} - \Delta\frac{\partial f}{\partial y}\Big|_{(\rho^2,(\rho^2-D),D)} + \frac{\Delta^2}{2}\frac{\partial^2 f}{\partial y^2}\Big|_{(\rho^2,y_0,D)} = \frac{1}{2}\ln\frac{\rho^2}{D} + \frac{\Delta^2}{2}\frac{\partial^2 f}{\partial y^2}\Big|_{(\rho^2,y_0,D)}, \tag{130}$$

where $\Delta = (\rho^2 - D)\bar{\alpha}$, and y_0 lies in the interval $(\frac{1}{2}(\rho^2 - D),(\rho^2 - D))$. Using (130) and the shorthand

$$f''(y_0) := \frac{\Delta^2}{2}\frac{\partial^2 f}{\partial y^2}\Big|_{(\rho^2,y_0,D)},$$

(128) can be written as

$$R\alpha - \frac{1}{2}\ln\frac{\rho^2}{D} = \bar{\alpha}^2\,\frac{(\rho^2-D)^2}{2}f''(y_0), \tag{131}$$

or

$$R - \frac{1}{2}\ln\frac{\rho^2}{D} = R\bar{\alpha}\ +\ \bar{\alpha}^2\,\frac{(\rho^2-D)^2}{2}f''(y_0). \tag{132}$$

Solving the quadratic in $\bar{\alpha}$, we get

$$\bar{\alpha} = \frac{-R + [R^2 + 2(\rho^2-D)^2(R - \frac{1}{2}\ln\frac{\rho^2}{D})f''(y_0)]^{1/2}}{(\rho^2-D)^2 f''(y_0)}. \tag{133}$$

Using this in (131), we get

$$\begin{aligned} &R\alpha - \frac{1}{2}\ln\frac{\rho^2}{D} \\ &= \frac{\left(-R + \left[R^2 + 2(\rho^2-D)^2(R - \frac{1}{2}\ln\frac{\rho^2}{D})f''(y_0)\right]^{1/2}\right)^2}{2(\rho^2-D)^2 f''(y_0)}. \end{aligned} \tag{134}$$

The LHS is exactly the quantity we want to bound from below. From the definition of f in (27), the second partial derivative with respect to y can be computed:

$$f''(y) = \frac{\partial^2 f}{\partial y^2}\Big|_{(\rho^2,y,D)} = \frac{\rho^2}{y^3} + \frac{1}{y^2} - \frac{y}{2(y^2+4\rho^2 D)^{3/2}}. \tag{135}$$

The RHS of (135) is strictly decreasing in y. We can therefore bound $f''(y_0)$ as

$$\begin{aligned} \frac{\rho^2}{(\rho^2-D)^3} < f''(\rho^2-D) < f''(y_0) < f''\left(\frac{\rho^2-D}{2}\right) \\ < \frac{12\rho^2}{(\rho^2-D)^3}. \end{aligned} \tag{136}$$

Substituting these bounds in (134), we conclude that for $D_\alpha \le D$,

$$\begin{aligned} g(D_\alpha) &= R\alpha - \frac{1}{2}\ln\frac{\rho^2}{D} \\ &\ge c_1 := \frac{(\rho^2-D)}{24\rho^2}\left(-R + \left[R^2 + \frac{2\rho^2(R - \frac{1}{2}\ln\frac{\rho^2}{D})}{(\rho^2-D)}\right]^{\frac{1}{2}}\right)^2. \end{aligned} \tag{137}$$

ACKNOWLEDGEMENT

We thank the anonymous referee for comments which helped improve the paper.

REFERENCES

[1] A. Barron and A. Joseph, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2541–2557, Feb 2012.

[2] A. Joseph and A. Barron, "Fast sparse superposition codes have exponentially small error probability for $R < C$," *IEEE Trans. Inf. Theory*, vol. 60, pp. 919–942, Feb 2014.

[3] R. Venkataramanan, A. Joseph, and S. Tatikonda, "Lossy compression via sparse linear regression: Performance under minimum-distance encoding," *IEEE Trans. Inf. Thy*, vol. 60, pp. 3254–3264, June 2014.

[4] R. Venkataramanan, T. Sarkar, and S. Tatikonda, "Lossy compression via sparse linear regression: Computationally efficient encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 60, pp. 3265–3278, June 2014.

[5] N. Alon and J. H. Spencer, *The probabilistic method*. John Wiley & Sons, 2004.

[6] M. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: Analysis and algorithms," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1351 –1368, 2010.

[7] S. Janson, *Random Graphs*. Wiley, 2000.

[8] A. Coja-Oghlan and L. Zdeborová, "The condensation transition in random hypergraph 2-coloring," in *Proc. 23rd Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 241–250, 2012.

[9] A. Coja-Oghlan and D. Vilenchik, "Chasing the k-colorability threshold," in *Proc. IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 380–389, 2013.

[10] A. Coja-Oghlan and K. Panagiotou, "Going after the k-SAT threshold," in *Proc. 45th Annual ACM Symposium on Theory of Computing*, pp. 705–714, 2013.

[11] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *Data Compression Conference*, pp. 53 –62, March 2011.

[12] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.

[13] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 20, pp. 197 – 199, Mar 1974.

[14] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with a fidelity criterion," *IEICE Trans. Fundamentals*, vol. E83-A, Oct. 2000.

[15] F. Den Hollander, *Large deviations*, vol. 14. Amer. Mathematical Society, 2008.

[16] R. R. Bahadur and R. R. Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, no. 4, 1960.